Next generation sequencing to understand norovirus in immunocompromised children

Thesis for the degree of

Doctor of Philosophy (PhD)

UCL

Julianne Rose Brown

2017

Declaration

I, Julianne Rose Brown, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Julianne Rose Brown

Abstract

Norovirus is a leading cause of gastroenteritis worldwide, causing self-limited vomiting and diarrhoea in immunocompetent people and chronic infections with significant morbidity in immunocompromised patients.

Data presented in this thesis uses deep sequencing to increase our understanding of norovirus in a hospital paediatric population with a large proportion of immunocompromised patients.

Real-time PCR reveals that norovirus is the most prevalent gastrointestinal virus in this population, causing infection with a higher viral titre than other gastrointestinal viruses. Norovirus is most common in immunocompromised patients and is the virus most commonly associated with chronic infections, which occur primarily in immunocompromised patients.

The performance of a novel method for deep sequencing norovirus full genomes is described; this overcomes the limitations of previously published methods and achieves full genomes with >12000-fold read depth regardless of genotype or viral titre. This method is applied to sequence the complete genomes of every new norovirus case at Great Ormond Street Hospital (GOSH) over a 19 month period. Full genomes reveal a broad range of circulating genotypes, more akin to genotypes circulating in the community than those typically seen in hospitals. Phylogenetic analysis shows that the majority (69%) of cases are not acquired from another patient. This suggests multiple introductions of different norovirus strains, with limited nosocomial transmission.

Full genome sequencing of longitudinally collected samples shows that chronic norovirus infections may involve super- or re-infection with a different genotype, although this does not affect the duration of infection.

Deep sequencing is used to investigate changes in the norovirus intra-host mutation frequency in chronically infected immunosuppressed patients who were and were not treated with oral ribavirin, revealing a possible role for ribavirin in the treatment of chronic norovirus infections. However interpretation of *in vivo* data is confounded by fluctuating mutation frequencies observed over time in all patients.

Table of Contents

List of Figures

List of Tables

Acknowledgements

I would like to extend my sincerest thanks to my supervisors, in particular Judy Breuer and Kathryn Harris, for their guidance, support and expertise throughout my PhD studies and in preparation of this thesis.

I am also grateful to Great Ormond Street Hospital (GOSH) Microbiology, Virology and Infection Control Department for hosting me during my doctoral fellowship and allowing me access to laboratory space, equipment and a unique collection of clinical specimens.

Due to the large number of specimens processed during the course of my doctoral studies, I have had help and support from Divya Shah, a research assistant at GOSH, and from the UCL Pathogen Genomics Unit (PGU) for some of the laboratory work. This support is detailed and acknowledged in each of the Chapters of this thesis, where applicable.

This work has been funded by a part-time National Institute of Health Research (NIHR) Healthcare Science Doctoral Research Fellowship (NIHR-HCS-D12-03-15), awarded to me in 2013. This fellowship has covered the cost of reagents and consumables, attendance at conferences and training courses, my salary and PhD fees.

In addition to my fellowship, the whole genome sequencing presented in this thesis was funded by Pathseek, an EU-funded collaborative grant headed by Professor Judy Breuer at UCL. The details of this support are given in Chapter 4.

Abbreviations

% OTR	percent on-target reads
µl	microliter
A	adenine
ANOVA	analysis of variance
B−	B cell deficient
B+	B cell non-deficient
BD	twice daily
BMT	bone marrow transplant
bp	base pair
C	cytosine
cDNA	complementary DNA
CS	clinical specialty
Ct	cycle threshold
CVID	common variable immunodeficiency
dN/dS	ratio of non-synonymous and synonymous nucleotide substitutions
DNA	deoxyribonucleic acid
EIA	enzyme immunoassay
EM	electron microscopy
FCV	feline calicivirus
FUT2	fucosyltransferase 2
G	guanine
Gb	gigabase
GOSH	Great Ormond Street Hospital
HBGA	human blood group antigen
HCV	Hepatitis C virus
HIE	human intestinal enteroids
HSCT	haematopoietic stem cell transplant
ICG	immunochromatographic
IPC	infection prevention and control
IVIG	Intravenous immunoglobulin

kb	kilo bases
kbp	kilo base pairs
mRNA	messenger RNA
NGS	next generation sequencing
NHS	National Health Service
NNUH	Norfolk and Norwich University Hospital
NPV	negative predictive value
NS7	norovirus non-structural protein 7 (RdRp)
nt	nucleotide
OD	once daily
ORF	open reading frame
P1	capsid protruding domain 1
P2	capsid protruding domain 2
PCR	polymerase chain reaction
PHE	Public Health England
PPV	positive predictive value
px	patient or patients
RBV	ribavirin
RCT	randomised controlled trial
RHDV	rabbit haemorrhagic disease virus
RdRp	RNA dependent RNA polymerase
RNA	ribonucleic acid
RNASeq	RNA sequencing
RT-qPCR	reverse transcription real-time PCR
SCID	severe combined immunodeficiency
SNP	single nucleotide polymorphism
T	thymine
TBE	Tris-Borate-EDTA
UTR	un-translated region
VLP	virus-like particle
VP1	norovirus structural protein 1 (major capsid protein)
VRD	Virus Reference Department

Relevant publications

The following articles have been published as a direct result of the work presented in this thesis:

Peer reviewed papers

Brown JR, Shah D and Breuer J (2016). Viral gastrointestinal infections and norovirus genotypes in a paediatric UK hospital, 2014–2015, *Journal of Clinical Virology*, **84**: 1–6.

Brown JR, Roy S, Ruis C, Yara Romero E, Shah D, Williams R and Breuer J (2016). Norovirus whole genome sequencing by SureSelect target enrichment: a robust and sensitive method, *Journal of Clinical Microbiology,* **54**(10): 2530–7.

Brown JR, Gilmour K and Breuer J (2016). Norovirus infections occur in B cell deficient patients, *Clinical Infectious Diseases*, **62** (9): 1136–8. **†**

† Chosen for editorial commentary, see: Green (2016). Editorial commentary: noroviruses and B cells, *Clinical Infectious Diseases*, **62** (9): 1136–8.

Book chapters

Brown JR and Breuer J (2016).Whole genome sequencing approach to genotyping and epidemiology, book chapter in *Norovirus: Features, Detection, and Prevention of Foodborne Pathogens*, edited by Dr. Paul Chan, Dr. H. S. Kwan, and Dr. Martin C. W. Chan. ISBN 9780128041772

Conference oral presentations

Brown JR, Roy S, Shah D, Ruis C, Williams R, Yara Romero E and Breuer J. Norovirus molecular epidemiology in a paediatric UK hospital: unexpected diversity, seasonality and sources of infection. Oral presentation at ESCV 2016, Lisbon (Portugal), September 2016.

Brown JR, Ruis C, Tutill H, Depledge D, Christiansen M and Breuer J. SureSelect target enrichment: A robust and sensitive method for sequencing of whole norovirus genomes direct from clinical specimens. Oral presentation at ESCV 2015, Edinburgh (UK), September 2015.

Lockwood JR*, Thorne L, Kundu S, Depledge D, Marks S, Fenton M, Goodfellow I and Breuer J. Potential use of oral ribavirin therapy for chronic norovirus infection in immunocompromised patients. Oral presentation at 5th International Conference Caliciviruses (presented by Lucy Thorne), Beijing (China), October 2013. *(maiden name of JR Brown)

Conference poster presentations

Lockwood JR*, Thorne L, Kundu S, Depledge D, Brown A, Holdstock J , Marks S, Fenton M, Goodfellow I and Breuer J. Oral Ribavirin Therapy for Chronic Norovirus Infection in Immunocompromised Patients. Poster presentation at ICAAC 2014, Washington (USA) September 2014. *(maiden name of JR Brown)

The following relevant publications were published before the commencement of work contributing to this thesis:

Kundu S*, **Lockwood JR*†**, Depledge D, Chaudhry Y, Aston A, Rao K, Hartley J, Goodfellow I and Breuer J (2013). Next generation whole genome sequencing identifies the direction of norovirus transmission in linked patients. *Clinical Infectious Diseases,* **57**(3): 407-414. *(maiden name of JR Brown) **†authors contributed equally**

Lockwood JR*, Cotten M, Goodfellow I, Kellam P and Breuer J. Development of a next generation sequencing pipeline for norovirus from clinical samples. Oral presentation at Noro2012: International Meeting on Norovirus and Other Caliciviruses on the Rise, Lübeck (Germany) March 2012. *(maiden name of JR Brown)

CHAPTER 1

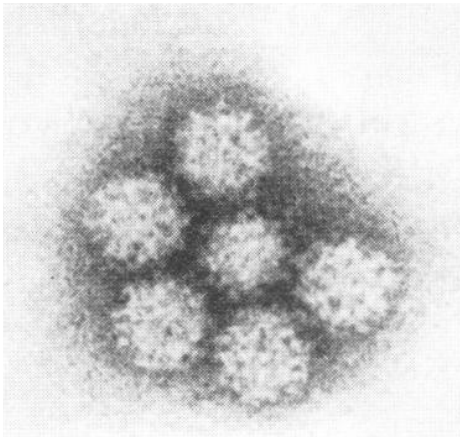
INTRODUCTION

1.1 NOROVIRUS BASIC VIROLOGY

1.1.1 Norovirus history and classification

Epidemic nonbacterial gastroenteritis was described as “Winter vomiting disease” in 1929 [1] however an aetiological agent could not be cultivated or identified. In 1972 immune electron microscopy, which involves the direct observation of antigen-antibody complexes, was for the first time used to analyse stool specimens obtained from an outbreak of acute gastroenteritis in Norwalk, Ohio (USA) that occurred in 1968. A small 27-nm particle with distinct surface morphology, including the now characteristic cup-like depressions on the surface of the virions (**Figure 1.1**), was identified. This virus was determined to be the cause of the outbreak by replicating symptoms in healthy volunteers experimentally infected with rectal swab filtrate taken from a patient linked to the 1968 outbreak [2]. The aetiological agent was termed Norwalk agent, and later Norwalk-like virus; in 2002 the genus was renamed as *Norovirus*, species *Norwalk virus* [3]. Norovirus was the first virus to be identified as a cause of diarrhoea.

Figure 1.1 First image of “Norwalk agent” by electron microscopy, 1972. Image reproduced, with permission, from Kapikian et al. (1972)

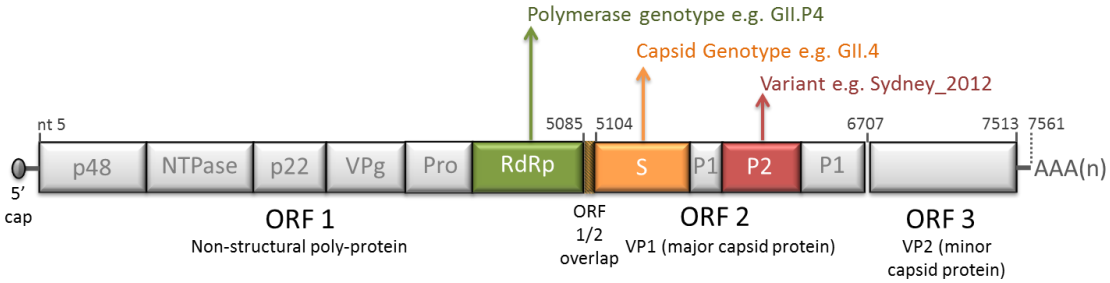


Noroviruses are a small non-enveloped virus with an icosahedral capsid measuring 27-40 nm in diameter belonging to the *Norovirus* genus in the family *Caliciviridae*. The *Caliciviridae* family comprises five phylogenetically distinct genera; *Norovirus, Sapovirus, Vesivivirus, Lagovirus* and *Nebovirus* [4]. The major human pathogens in this family are the norovirus and sapoviruses; other important pathogens in the family are veterinary pathogens, namely feline calicivirus (FCV) in the *Vesivirus* genus and rabbit haemorrhagic disease (RHDV) in the *Lagovirus* genus. FCV and RHDV cause respiratory disease in cats and haemorrhagic disease in rabbits, respectively.

1.1.2 Genome and structure

Noroviruses have a linear, positive-sense single-stranded RNA genome. The 7.5 kb genome, first described by Jiang *et al.* in 1990 and characterised in 1993 [5, 6], comprises three open reading frames (ORF) with a polyadenylated 3' end (**Figure 1.2**). The 5' end of the genome is covalently attached to a virus-encoded protein, VPg (5' cap), which recruits host cell transcription initiation factors. The un-translated regions (UTRs) are typically very short, with 5 and 48 nucleotides at the 5' and 3' ends, respectively.

Figure 1.2 Schematic of norovirus genome, with genotyping regions highlighted. Nucleotide numbering is based on a GII.4 genome (p48, replication complex formation; NTPase, NTP binding and hydrolysis of NTP; p22, replication complex formation; 5' cap, viral genome-linked protein (VPg)involved in initiation of translation; Pro, cysteine proteinase, poly-protein cleavage; RdRp, RNA dependent RNA polymerase; S, shell domain; P1, protruding domain 1; P2, protruding domain 2).



ORF 1 codes for a non-structural polyprotein which is cleaved post-translation by a virus-encoded proteinase (Pro) into six non-structural proteins involved in viral replication, including the RNA dependant RNA polymerase (RdRp) [7]. ORFs 2 and 3 are translated from sub-genomic RNA produced during virus replication; ORF 2 codes for the major capsid protein (VP1) and ORF 3 for a minor capsid protein (VP2). The role of VP2 is poorly understood, but is most likely involved in capsid assembly and genome encapsidation [8]. Murine norovirus is the only norovirus that has a fourth ORF, overlapping the 5' end of ORF 2, which codes for a protein involved in the regulation of the innate immune response in mice.

The RNA-dependent RNA polymerase (RdRp) protein has a "right hand" structure with finger, thumb and palm domains. The finger and palm domains, bridged by an N-terminal domain, form a rigid unit; the thumb domain is flexible, assuming either a "closed" or "open" conformation. The C-terminal end of the protein lies within the active-site cleft when the polymerase is not bound to a template. The C-terminal tail is

displaced from the active site to allow binding of the primer–template RNA duplex and thus initiation of RNA synthesis [4].

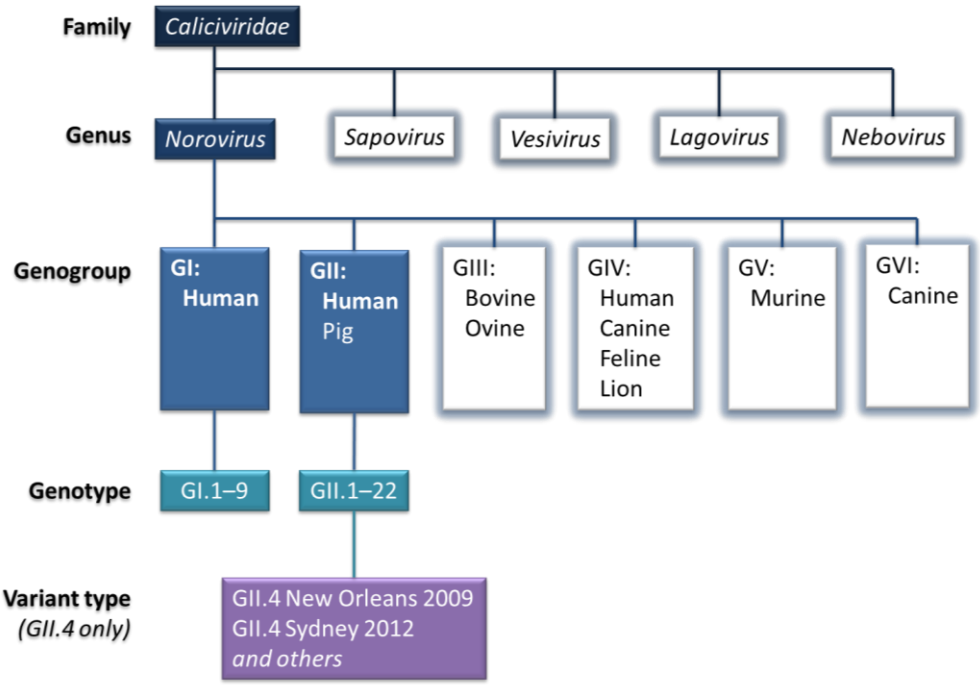
The norovirus capsid consists of the N-terminal shell (S) domain and the protruding (P) domain. The S domain surrounds the viral RNA whilst the P domain, linked to the S domain through a flexible hinge, is the surface-exposed region of the capsid. The P domain has two subdomains, P1 and P2. P2 is the most exposed antigenic site and a receptor-binding region; consequently it has the greatest sequence variation.

The major capsid protein (VP1) can self-assemble into virus-like particles (VLPs) without the requirement of RNA or VP2. This allows recombinant VLPs to be expressed in baculovirus systems, thus serving as a surrogate for full virions in the absence of a cell culture system.

1.1.3 Nomenclature

The *Norovirus* genus is composed of six genogroups (GI to GVI) based on genetic relatedness (**Figure 1.3**); GI and GII, and to a lesser extent GIV, cause infections in humans. The remaining genogroups cause diarrhoeal infections in pigs, cows, sheep, cats, dogs and rats, and sub-clinical infections in mice [9-15]; however there is no evidence of zoonotic transmission of animal noroviruses to humans [16, 17]. Genome sequences between the genogroups are highly diverse with, on average, 32-35% amino acid differences between genotype capsid sequences [18].

Figure 1.3 Norovirus nomenclature. Variant types only apply to GII.4 noroviruses



Genogroups GI and GII are categorised into 9 and 22 genotypes, respectively (GI.1–GI.9 and GII.1–GII.22). Due to recombination between genotypes at the ORF1/ORF2 junction, norovirus has a dual typing system based on the polymerase (ORF1) and capsid (ORF2) sequences. Thus a norovirus with a genogroup II, genotype 4 polymerase and capsid type will be designated GII.P4_GII.4, whereas a recombinant sequence with a GII.3 polymerase and a GII.4 capsid type will be designated GII.P3_GII.4. Orphan polymerase sequences, which are novel polymerase types that have only been seen in combination with established ORF2 genotypes instead of a unique ORF2 genotype, are designated a letter such as GII.Pc or GII.Pe, for example GII.Pe_GII.4. Despite the occurrence of recombination and the recommendation of a dual typing system involving both capsid and polymerase sequences [18], norovirus genotyping is routinely often still based on capsid typing alone.

Diversity among noroviruses is maintained through two principle mechanisms. The first is recombination between norovirus genotypes, with a break-point at the ORF1–ORF2 junction [19, 20]. The second, which is responsible for within-genotype diversity, is the accumulation of point mutations associated with error-prone replication by the RNA-dependent RNA polymerase, which lacks a proof-reading mechanism [21]. The error-prone template copying by RNA polymerase can lead to point mutations, either silent or resulting in amino acid substitutions, which in turn can lead to genotype diversity, antigenic variation and immune escape [22].

1.1.4 Norovirus biology

Understanding of the norovirus life cycle has been limited due to the inability to culture human norovirus *in vitro* [23]. In 2014, Jones *et al.* published the first report of successful cultivation of human norovirus in a B cell line [24] however this was not replicated by other laboratories [25] therefore the challenge remained.

Based on clinical features, the primary site of human norovirus infection and replication is assumed to be the small intestine; however cell types infected by norovirus are not defined. A failure to cultivate norovirus in transformed epithelial cell lines [23, 26-30] and a failure to observe viral particles in enterocytes from intestinal biopsies from infected volunteers [31, 32] suggested enterocytes were not the primary site of norovirus infection and replication. However, histopathological analysis of intestinal biopsies from chronically infected immunocompromised patients demonstrated the presence of structural and non-structural viral proteins in duodenal and jejunal enterocytes [33], suggesting enterocytes may be a permissive cell type for human norovirus replication after all, perhaps necessitating a primary culture system derived

from intestinal cells. The reason why norovirus was not originally detected in the enterocytes of infected volunteers [31, 32] is unclear; it may be due to differences in the primary site of replication between acute and chronically infected patients. However the original studies were conducted in 1973 and 1975 using electron microscopy, whereas the study reporting norovirus proteins in the enterocytes of a chronically infected patient [33] was conducted in 2016 using antigen-specific immunofluorescence to detect viral proteins. Therefore it is plausible that the discrepancy in the studies' results is attributable to differences in the sensitivity of the detection methods employed, rather than differences in the site of replication.

In 2016 Ettayebi *et al.* successfully cultured human norovirus in an intestinal primary culture system, known as human intestinal enteroids (HIEs) [34]. HIEs are generated from stem cells isolated from intestinal crypts in human intestinal tissues. They are multicellular, differentiated, non-transformed physiologically active cultures that recapitulate the natural intestinal epithelium. HIEs contain multiple intestinal epithelial cells; namely enterocytes, goblet, enteroendocrine and Paneth cells. They exist as either three dimensional or monolayer cultures. Ettayebi *et al.* successfully infected HIE monolayers with human norovirus GII.4, resulting in a 2.5 log increase in viral copies, cytopathic effect (CPE) with cell rounding and destruction of the monolayer and progeny virus that could be passaged. The authors demonstrated that despite cultures containing multiple cells types only enterocytes were infected, leading to the suggestion that enterocytes are the likely primary site of replication *in vivo*. HIEs were also successfully infected with GII.3, GII.17 and GI.I norovirus strains, however with lower levels of replication. Replication of non-GII.4 strains required the addition of bile, which was essential for virus adsorption. The active component in bile has not yet been characterised, however it is expected to be non-proteinaceous as heat or trypsin treatment did not alter its activity.

Noroviruses are known to recognize and interact with human blood group antigens (HBGAs) [35], which are carbohydrate epitopes present on the surface of red blood cells, mucosal epithelia or as free antigens in biological fluids such as saliva and intestinal contents [36]. The human population can be divided into two broad groups; "secretors" and "non-secretors". An estimated 70-80% of the human population are considered "secretors" [37], in whom *fucosyltransferase 2* (*FUT2*) gene expression controls the expression of HBGA on mucosal epithelium, including in the gut. In the remaining 20–30%, referred to as "non-secretors", *FUT2* expression is inactivated by a single nucleotide polymorphism (SNP), therefore there is no HBGA expression on the intestinal epithelium. Challenge studies have demonstrated that *FUT2* non-secretors

are inherently resistant to infection with GI.1 and GII.4 noroviruses [38-40], which is backed up by epidemiological studies in which infection with GII.4 noroviruses is only observed in secretors [41]. However non-secretor status is not protective against non-GII.4 norovirus infections [42, 43]; this extends to *in vitro* infection, for which Ettayebi *et al.* demonstrated that whilst GII.4 norovirus cannot infect HIEs generated from non-secretors, GII.3 infected HIEs from secretor and non-secretors. It is already known that the specific HBGAs to which norovirus binds varies with genotype, with epidemic GII.4 noroviruses binding A, B and O antigens [44], however the full extent to which secretor or non-secretor status protects against infection with different genotypes is not known.

HBGAs, as well as being expressed in human hosts, are expressed by certain bacteria [45-47]. Jones *et al.* reported that *in vitro* infection of a B cell line was only enabled by the addition of enteric bacteria expressing HBGA [24]; as such norovirus infection of the cell line could only be achieved from unfiltered stool that contained bacteria or with the addition of synthetic H-type HBGA. The role of enteric bacteria was further confirmed *in vivo* in murine infections, which demonstrated reduced norovirus titres in mice treated with antibiotics to deplete their intestinal microbiota [48]. These data suggests the role of HBGA binding, rather than direct viral attachment to the intestinal wall, may be to attach to enteric bacteria which stimulates infection of permissive cells [49]. However in Ettayebi *et al.*'s HIE primary culture system, infection was achieved using norovirus positive stool filtrates, which do not contain HBGA-expressing bacteria [34]. Therefore the precise role of HBGA, enteric bacteria and co-factors for norovirus infection remain incompletely defined.

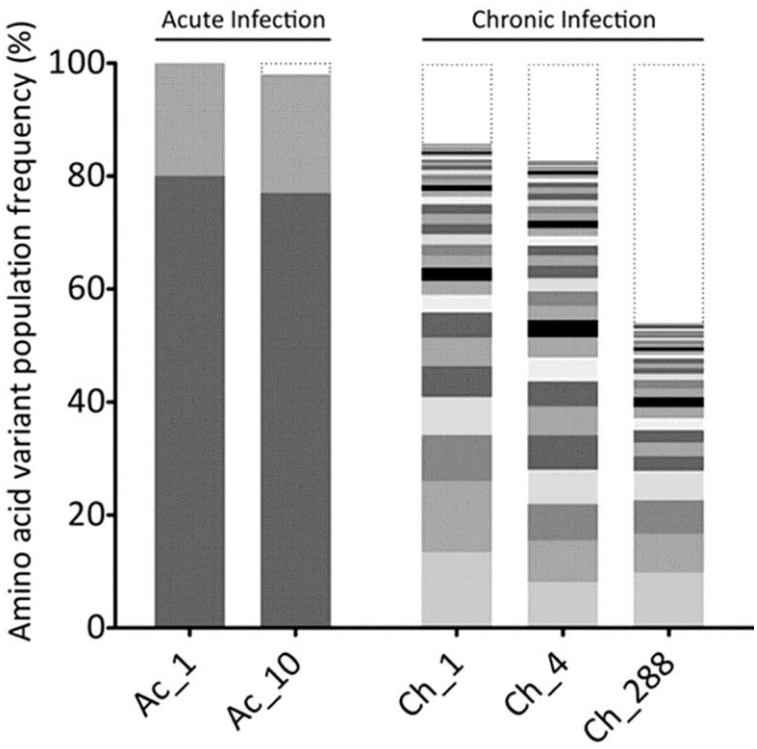
Prior to the recent advances in human norovirus culture, the only norovirus that could be cultured *in vitro* was murine norovirus. Consequently the mouse model has been used to study norovirus biology and host responses to infection. However murine norovirus infection in immunocompetent mice causes a sub-clinical persistent infection without diarrhoea therefore is not truly representative of human norovirus infections.

In the absence of an *in vitro* culture system, some human norovirus replication studies have been possible through the use of replicon bearing cells. RNA transcripts are derived from plasmid DNA containing a full length norovirus genome; these transcripts are then transfected into a stable cell line, creating replicon-bearing cells. Replicon-bearing cells can maintain the RNA after multiple passages and express norovirus proteins, with the exception of the capsid structural protein VP1 [50].

Replication in RNA viruses is characteristically error-prone, caused by a lack of proof-reading mechanisms that result in limited replication fidelity. Consequently intra-host

viral populations exist as a quasispecies, with a degree of sequence heterogeneity within a host [51]. In immunocompetent individuals, the norovirus intra-host viral population is relatively homogeneous. Conversely, in immunodeficient individuals with chronic infection the intra-host viral population is highly heterogeneous, with no single dominant strain (**Figure 1.4**) [52]. The high diversity in these patients has led to the suggestion they may be a reservoir for the emergence of novel viral variants [52, 53].

Figure 1.4 Comparison of the intra-host distributions of norovirus variants in immunocompetent and immunocompromised subjects, determined by deep sequencing ORF2 (figure reproduced, with permission, from Bull et al. 2012 [52]). Each unique variant is represented by alternate grey shading; number on x-axis indicates day of sampling. Low-frequency variants (<2%) are indicated by black dotted lines. For the immunocompetent subject (Ac), only two variants were detected, with frequencies of 79% and 20%, respectively. These variants remained stable over the 9 days of infection. For the immunocompromised subject (Ch), no dominant variant was observed. Instead, a distribution of low-frequency variants coexisted, and their prevalence varied over the course of the infection.



1.1.5 Immunology

Since norovirus infections in immunocompetent people typically resolve within days, the innate immune response is predicted to play a critical role in controlling infection. Nevertheless there is evidence to suggest that the adaptive and cell-mediated immune response also contributes to control of infection, including neutralizing antibodies, CD4+ T cells and CD8+ T cells [49]. The role of T cells in the control of infection has

been demonstrated in both animal models and human case reports, in which norovirus clearance in immunocompromised patients is associated with immune reconstitution and T cell recovery [54, 55]. Furthermore, in volunteer studies that involved oral immunization with recombinant virus-like particles (VLPs) the VLPs elicited a virus-specific T-helper type 1 (Th-1)-dominated cell-mediated response [56].

Norovirus infection elicits mucosal and serum antibody responses, which in mouse models are protective against secondary norovirus infections [57, 58]. Furthermore, in human infections norovirus-specific mucosal and serum IgA correlates with protection [38, 59]. However, early human volunteer studies did not demonstrate cross-protective or long term protective immunity following infection [40, 60]. Infected volunteers were resistant to re-infection 6–14 weeks after the initial challenge, but only if re-challenged with the same genotype [61]. Two explanations are offered for the observed lack of protection. First, noroviruses are genetically diverse, with weak to no cross-protective immunity across genotypes. Second, protective immune responses are short-lived. Estimates from mathematical modelling suggest immunity to human norovirus lasts 4–9 years [62], however a study showed that whilst volunteers were protected against homologous re-infection 6 months after initial exposure, immunity had waned by two years [60].

1.2 NOROVIRUS INFECTIONS

1.2.1 Clinical presentation

Norovirus transmission occurs via the faecal-oral or vomit-oral route, either directly person-person (88% of 5036 outbreaks in Europe, 2001–2006), via faecally contaminated food or food contaminated by infected food handlers (10%) or contaminated water (1.5%) [63]. Transmission can also occur indirectly via contaminated surfaces or fomites [64]. Norovirus has a low infectious dose, estimated at only 18 virus particles [65].

Following a 12-36 hour incubation period, norovirus infections are typically associated with an acute onset, projectile vomiting lasting less than a day and watery diarrhoea lasting two to three days. In immunocompetent people infections are generally self-limiting with limited morbidity besides dehydration.

Virus shedding persists after resolution of symptoms; on average norovirus is detected in stool for up to 22.1 days in infants under one year old and 14.5 days (range 2–34) in immunocompetent adults and children over one year of age [66]. In the absence of an *in vitro* culture system or animal model for human norovirus, it is not known whether

virus shed following the resolution of symptoms is viable and therefore infectious. Although transmission is typically caused by symptomatic, rather than asymptomatic patients [67], there is no direct evidence to suggest asymptomatically shed virus is non-infectious. It is possible that limited transmission following the resolution of symptoms is due to reduced faecal contamination of the environment in the absence of diarrhoea, as opposed to shed virus being non-viable; evidence of infectivity is required to confirm this.

The pathological basis of diarrhoea in norovirus infections is not well understood, however infections are associated with broadening and blunting of villi and transient malabsorption of D-xylose, fat and lactose. The cause of diarrhoea is postulated to be an alteration of secretory and/or absorptive processes, rather than structural damage of the intestinal wall [49]. Typically there is only moderate intestinal inflammation, although severe infections may be associated with higher gut inflammation [68]. The underlying pathophysiology of vomiting is undefined.

Based on a single outbreak study, norovirus is suggested to cause post-infectious irritable bowel syndrome in 34% of infected patients; however this is transient, having resolved by 6 months post-infection [69]. In patients with inflammatory bowel disease, norovirus infection can present with bloody diarrhoea [70]. In rare instances, norovirus has been reported as the cause of necrotizing enterocolitis of the large and small intestine in premature infants [71-73], benign infantile seizures [74, 75] and encephalopathy [76, 77].

In immunocompromised patients, norovirus infection can cause a bi-phasic illness and is increasingly recognised as a significant cause of morbidity [78, 79]. The initial acute phase exhibits classical symptoms of vomiting and diarrhoea, however a second chronic phase may ensue with viral shedding and diarrhoea lasting weeks to years. Long-term shedding in immunocompromised patients in the literature ranges from 35-898 days [66]. The consequence of chronic infection can be dehydration, malnutrition, dysfunction of intestinal barrier [80], dramatic weight loss [81] and a requirement for nutritional support [54]. Deaths attributed to chronic norovirus infections are rare but have been reported in immunocompromised patients [81, 82]. Since norovirus clearance in immunocompromised patients is associated with immune reconstitution, cautious reduction of immunosuppression in chronically infected solid organ transplant patients may contribute to viral clearance, however this risks graft rejection or graft versus host disease (GHVD). In the absence of full immune reconstitution there is a need for specific antiviral therapy to clear chronic norovirus infections and recover gut function.

Norovirus can also cause asymptomatic infections, in which viral genomes are detected in stool but there are no accompanying gastrointestinal symptoms. The overall prevalence of sub-clinical infections in England is estimated to be 12% across all age groups, with the greatest burden of asymptomatic infection falling to under five-year olds, in whom prevalence exceeds 25% [83]. In a small cohort of children in France with inherited immunodeficiencies 11.6% (5/43) of asymptomatic children had detectable norovirus in stool, thus were presumed to have sub-clinical infections [84]. Asymptomatic patients are less likely than symptomatic patients to transmit norovirus to others [67], presumably due to reduced environmental spread in the absence of vomiting or diarrhoea. The viral load and duration of norovirus shedding in stool has been shown to be comparable in symptomatic and asymptomatic cases [85].

1.2.2 Diagnosis of infection

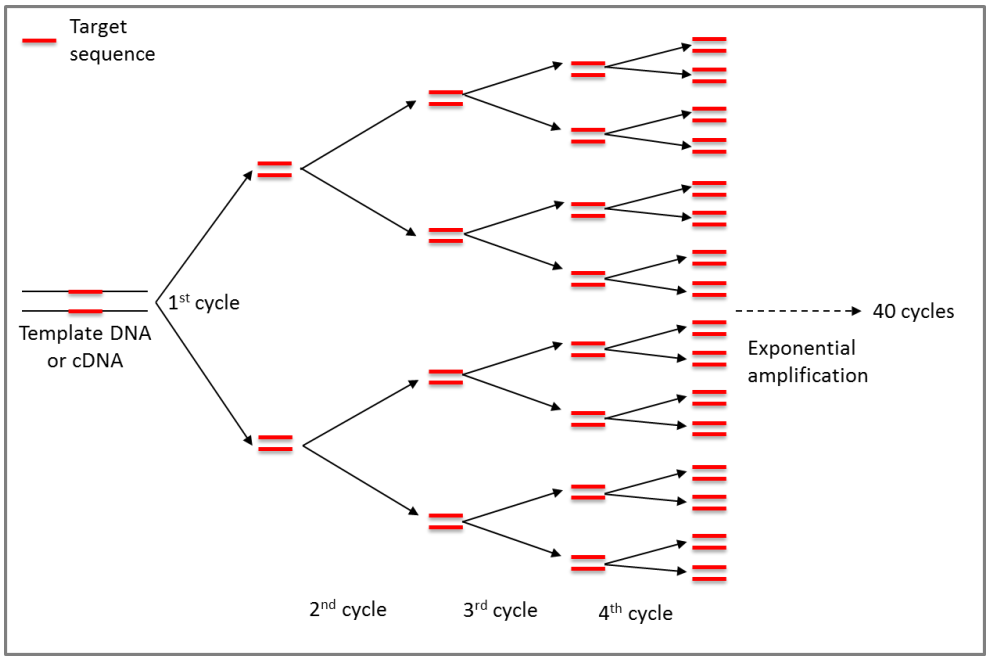
Before the advent of molecular techniques in diagnostic settings, norovirus diagnosis relied on electron microscopy of stool. However this lacked both specificity and sensitivity; structurally similar gastroenteric viruses, such as *Sapovirus* also in the *Caliciviridae* family, could be misidentified as norovirus. In addition a tendency for norovirus virions to clump together means they are not homogenously distributed in stool therefore may be missed. Consequently electron microscopy is no longer widely used in diagnostic laboratories.

Immunochromatographic (ICG) lateral flow tests can provide a rapid result (15 minutes) and be performed by persons without technical or laboratory expertise, therefore have a potential role in bed-side diagnostics. However, the sensitivity of these tests has been shown to be very poor, only 35–52%, and highly genogroup dependent [86]. The sensitivity of enzyme immunoassays (EIA) to detect norovirus antigen is also reportedly very low (<70%) therefore they are not recommended for diagnosis of norovirus infection [87].

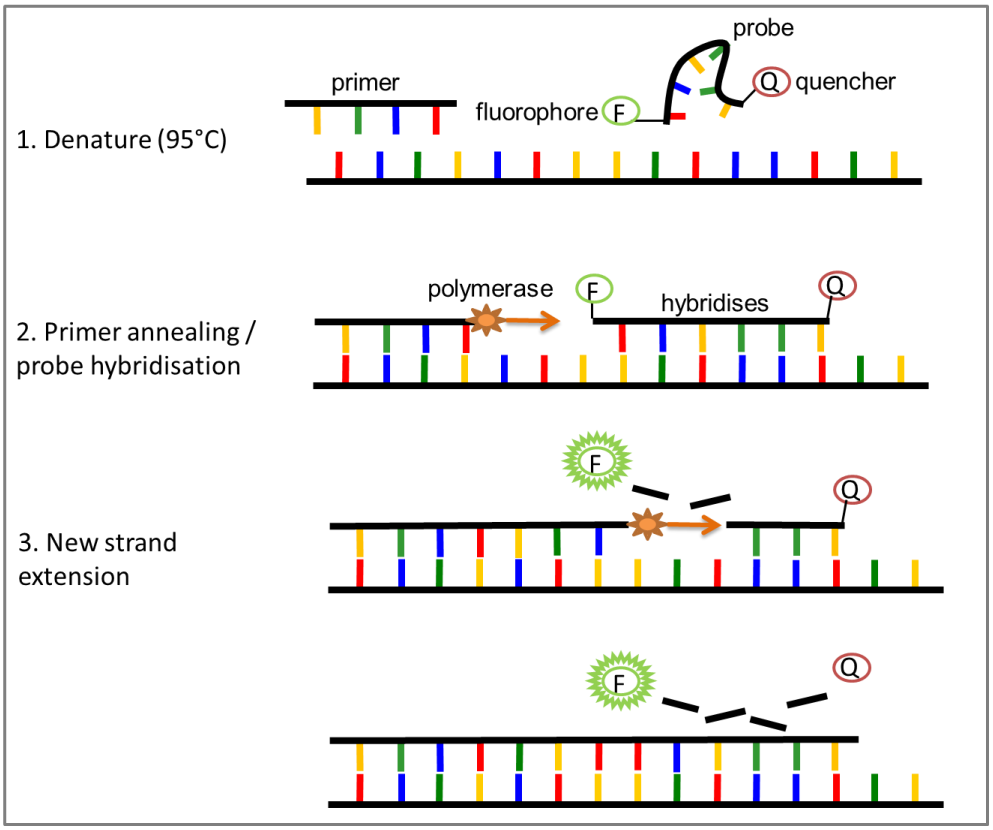
The gold standard of diagnosis for norovirus and other viral infections in the majority of diagnostic laboratories is reverse-transcription real-time PCR (RT-qPCR). Real-time PCR uses primers and a fluorescently-labelled probe specifically targeting the virus of interest, in this case norovirus. The fluorescent signal emitted by the probe increases exponentially with amplification of the target DNA or cDNA (**Figure 1.5**).

Figure 1.5 (a) Amplification and **(b)** fluorescent signal emission in a real-time PCR reaction. F, fluorophore; Q, quencher. Separation of the fluorophore and quencher during probe hydrolysis emits a fluorescent signal, which is detected during each cycle of PCR

(a)

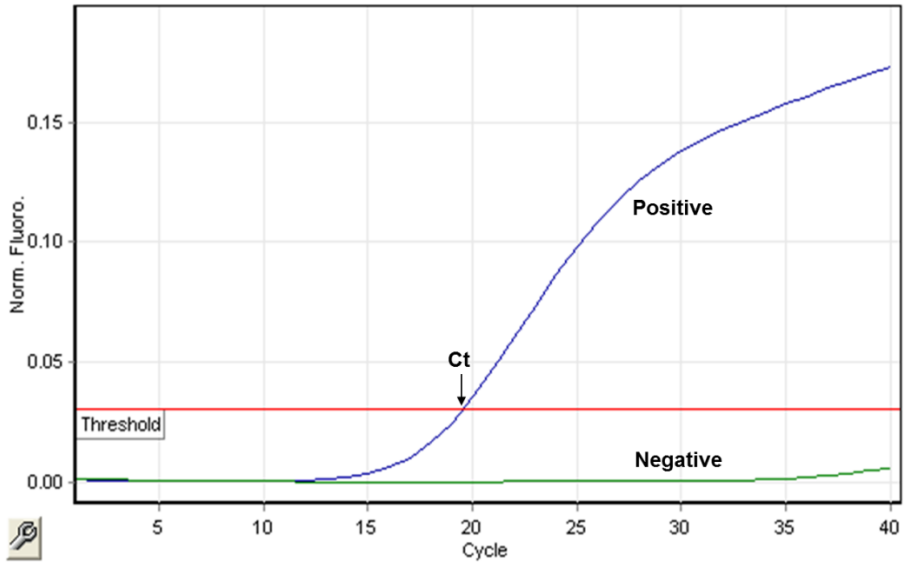


(b)



The PCR cycle at which the fluorescent signal crosses a pre-defined threshold (set above background noise) is known as the Cycle Threshold (Ct); this is the semi-quantitative value with which we report a positive result (**Figure 1.6**). High viral loads will result in an amplification signal crossing the threshold earlier in the PCR cycles, thus an earlier Ct value, and vice-versa. There is a linear relationship between viral load and Ct value; a difference in 3 Ct values is equivalent to an approximate 10-fold difference in viral load, and a difference in 1 Ct value approximately a two-fold difference in viral load.

Figure 1.6 Positive (detection) and negative (no detection) real-time PCR amplification curves



1.2.3 Control of infection in hospitals

In the absence of a vaccine or therapeutic options, control of norovirus outbreaks in hospital is achieved with strict infection control practices to interrupt transmission, as recommended by the Public Health England (PHE) guidance for managing norovirus outbreaks in acute and community health and social care settings [88].

Good hand hygiene should be actively encouraged in patients, staff and visitors. Due to a lack of viral envelope, norovirus is resistant to common disinfectants, including alcohol-based hand sanitizers [89-92]; therefore hand washing must include soap and running water.

Norovirus can be detected in the majority of vomit samples from infected patients [64] and droplets and spray from projectile vomiting can reach a 7.8 m^2 area, therefore vomiting is likely to be an important mode of norovirus transmission. Given that norovirus remains stable in the environment for weeks to months [93], daily environmental cleaning with a bleach solution, followed by a terminal clean after patient

discharge or after resolution of symptoms, is an important aspect to control and prevent the spread of norovirus [94]. The difficulty in eliminating norovirus from the environment was demonstrated by Morter *et al.*, who detected norovirus by PCR in 31% of environmental swabs after cleaning bays or wards that had accommodated norovirus infected patients; norovirus was detected in the environment surrounding the bedside, furniture, toilet and shower room fixtures and fittings as well as notes trolleys, keyboards and soap and alcohol dispensers [95].

Control of norovirus outbreaks is more difficult in hospitals with nightingale-style wards, as these are often associated with high patient density, limited hand washing facilities and minimal or no barriers between patients. Fewer norovirus outbreaks are observed in hospitals with small bays or a high proportion of side-rooms for patient isolation compared to those with nightingale-style wards and limited isolation facilities [96].

Isolation and cohorting is a major part of norovirus infection control practices; this includes exclusion, isolation or cohorting of infected patients and exclusion of infected staff for 48 hours after symptom resolution [88]. Exclusion of infected staff members can reduce transmission to patients [97], however exclusion can also involve excluding new admissions onto wards affected by a norovirus outbreak by closing the ward; in 2015 PHE reported bay or ward closures in 94% (611/653) of nosocomial outbreaks [98]. Closure of affected hospital wards to new patient admissions during a norovirus outbreak can be an effective method to limit transmission; most data suggest that ward closures are effective in terms of the number of cases and duration of outbreak [94]. In Avon, UK, the closure of units within four days resulted in outbreak containment within one week, compared to two weeks without rapid unit closure; however this came at a cost of £635,000 per 1,000 beds [99]. Owing to the cost and disruption to patient care associated with ward closure, UK guidelines are shifting away from ward closure recommendations [88]. A recent UK analysis of ward closure cost effectiveness in norovirus outbreaks concluded that whilst ward closures are highly likely to avert cases of nosocomial infections, it is also highly likely to increase costs to the hospital; there is therefore a very low probability that ward closure saves costs. Closing a ward on the third day of an outbreak will result in a third fewer subsequent cases and outbreaks but is likely to incur a loss of £14,000 [100]. However if a hospital is willing to pay £5,000 to avoid an additional nosocomial case then, assuming ≥25% efficacy at preventing transmission to other wards, ward closure becomes cost-effective. For ward closure to be cost-effective in preventing an outbreak the hospital must be willing to pay £35,000 and, additionally, the efficacy of the ward closure must exceed 50% [100].

Norovirus vaccination studies for the prevention of norovirus infection are underway [101]. Phase I and II norovirus vaccine trials, using non-replicating virus-like particles (VLP), have shown promise for the prevention of norovirus infection; the incidence and disease severity is significantly reduced in vaccinated compared to placebo groups [102, 103]. However re-infection challenges were only within weeks of vaccination, therefore the duration of protection is unclear. Moreover, human vaccine studies which included a re-infection challenge did not involve a heterologous genotype, therefore cross-genotype protection is unknown [101].

1.2.4 Disease burden

Diarrhoeal disease constitutes the second greatest burden of all infectious diseases in respect of disability adjusted life years and deaths worldwide each year, second only to lower respiratory infections and preceding HIV [104, 105], causing an estimated four episodes of diarrhoea per year in a child under 5 years in Europe [106]. With an estimated prevalence of 20% in cases of acute gastroenteritis in developed countries [107] norovirus is considered the leading cause of outbreaks of acute gastroenteritis [108, 109]. In countries where rotavirus vaccine has been introduced, norovirus is now the most common cause of medically-attended gastroenteritis in children [110, 111]. The majority of the global population will encounter norovirus before adulthood [4]; in a study of Finnish children aged 0–14 years, antibody prevalence against norovirus GII.4 reached 91.2% after 5 years of age [112].

The annual cost to the NHS of infectious gastroenteritis from bed days lost and staff absences is estimated at £115 million, of which 63% is attributable to norovirus outbreaks [99]. The global economic burden of norovirus has recently been estimated at US$4.2 billion in direct healthcare costs. However since the majority of infections are short and self-limiting without need to present to healthcare services, norovirus infections seen in healthcare institutions are only the tip of the iceberg. For every norovirus case reported to national surveillance, there are an estimated 147 additional cases in the community which are unreported [113]. Consequently the global societal costs from norovirus infections, primarily attributed to loss of productivity, are estimated at US$60.3 billion per year, with disease in under 5-year olds accounting for the biggest portion of economic burden (US$39.8 billion) [114].

Diarrhoea is a common complication in transplant recipients, with up to 80% of hematopoietic stem cell transplant (HSCT) patients suffering from gastroenteritis [115]. There are several possible causes, including conditioning therapy, graft versus host disease (GVHD), drugs or an infectious agent. The prevalence of norovirus infection in

immunocompromised patients has not been determined; however in case series of immunocompromised patients it has been shown to be 17-23% [54, 116-118].

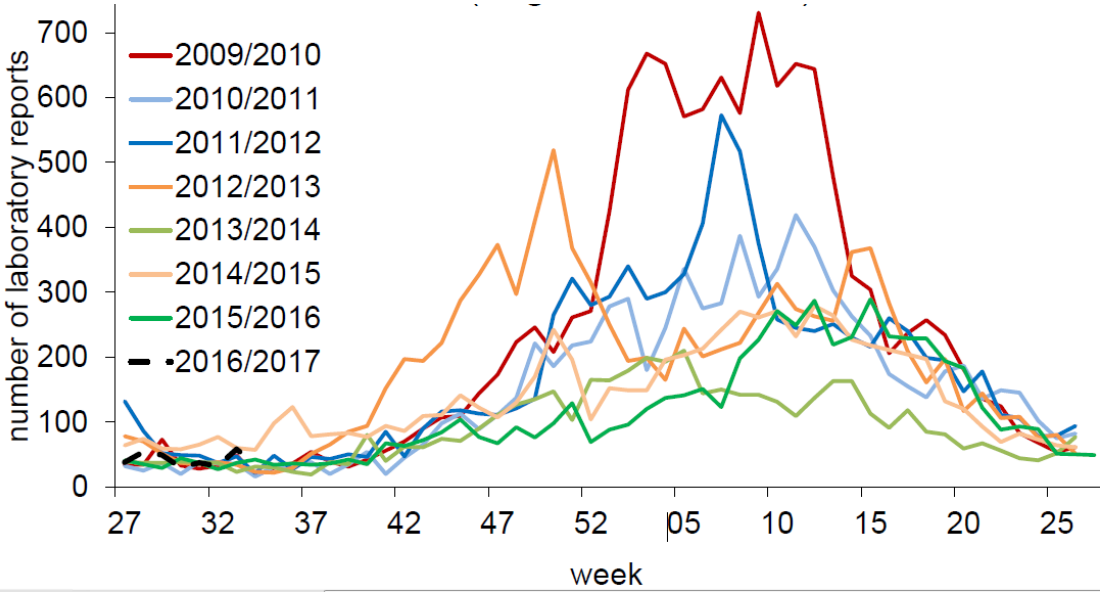
1.2.5 Norovirus epidemiology

1.2.5.1 Global epidemiology

Norovirus causes sporadic cases of gastroenteritis in the community, occurring in individuals or small family clusters. Causing an estimated 3 million sporadic cases of gastroenteritis annually in the UK [113], it is estimated that for every case of norovirus reported to national surveillance there are 147 community cases and 10 general practice (GP) consultations [113]. Sporadic community cases of norovirus occur year-round with no distinct seasonality [119], as is the case for infections in immunocompromised hosts [82, 120].

In addition to sporadic infections, norovirus is characterised by outbreaks occurring each year with a winter-time peak [121] (**Figure 1.7**); for this reason norovirus is often referred to as "winter vomiting disease". Norovirus outbreaks typically occur in semi-enclosed settings such as hospitals, long-term care facilities, schools, military bases and cruise ships [122].

Figure 1.7 Seasonal comparison of laboratory reports of norovirus (England and Wales). 2009/2010 and 2012/2013 are epidemic years with heightened and early seasons, respectively, caused by the emergence of novel GII.4 variants. Figure reproduced from PHE national norovirus surveillance report available at https://www.gov.uk/government/statistics/norovirus-national-update (accessed 22/09/2016)



Norovirus epidemiology is punctuated by epidemic years, in which there is heightened norovirus activity or an increase in cases outside of the typical winter season, as observed in 2009/2010 and 2012/2013, respectively, in **Figure 1.7**. These are caused by the emergence and rapid global spread of a novel GII.4 variant, which replaces the previously dominant variant. Since 1995 norovirus pandemics have occurred every 2–3 years, as listed in **Table 1.1**. The current circulating GII.4 variant worldwide is Sydney_2012, which replaced New Orleans_2009 in 2012–2013.

Table 1.1 Norovirus GII.4 epidemic variants

Variant Name	Year first emerged
US95_96	1995
Farmington_Hills_2002	2002
Asia_2003	2003
Hunter_2004	2004
Yerseke_2006a	2006
Den Haag_2006b	2006
New Orleans_2009	2009
Sydney_2012	2012

The emergence of novel GII.4 variants is associated with amino acid residue changes in three surface-exposed epitopes of the hyper-variable capsid P2 domain, listed in **Table 1.2**. Changes in these epitopes are predicted to create antibody-escape mutants to which the population has little or no cross-protective immunity. The ability of pre-existing immunity to neutralize the novel virus is limited, resulting in emergence of a new pandemic strain [123, 124]. Homologous substitutions at these immunogenic sites do not change the properties of virus binding or immunogenicity of the virus epitopes [125].

Table 1.2 Norovirus immunogenic epitopes associated with shifts in GII.4 antigenicity and emergence of new epidemic strains.

Epitope	Capsid amino acids involved
Epitope A [123]	294, 296–298 (Site A)*, 368, 372
Epitope D [124]	393–395 (Site B)*
Epitope E [124]	407, 412, 413

* Site A and Site B initially described by Allen *et al.* (2008) [125]; superseded by Epitope A and D in later publications

As part of national outbreak control in the UK, Public Health England (PHE) undertakes continuous surveillance of norovirus activity via the following means [88]: Calls to the

National Health Service (NHS) telephone service, NHS Direct, are monitored as an early warning sign; an increase in vomiting in the community can indicate the beginning of the norovirus “season” and can give a four week warning of impending norovirus pressures on the health service [126]. Voluntary weekly reporting of suspected and laboratory-confirmed norovirus hospital outbreaks to PHE via the online Hospital Norovirus Outbreak Reporting Tool (http://bioinformatics.phe.org.uk/noroOBK) to identify norovirus epidemics in a timely manner. Lastly, surveillance of norovirus outbreak strains to identify the emergence of new strains, which is a predictor of the impact of outbreaks.

1.2.5.2 Circulating genotypes

Since the mid-1990s norovirus outbreaks globally have been dominated by the genotype GII.4 [127], implicated in more than 90% of outbreaks [128]. The genotypes causing the non-GII.4 proportion of outbreaks vary between season, with GII.2, GII.3, GII.6 and GII.7 detected most frequently [129, 130].

For the first time in almost 20 years, GII.4 has been outcompeted by an emerging genotype, GII.17 Kawasaki 2014, in China and Japan [131]. During the winter of 2014–2015 GII.17 caused 66% of hospitalized norovirus cases in Hong Kong, compared to 19% caused by GII.4. GII.17 reportedly causes more infections in elderly people, causing infection in only 16% of children under five and in 37% of over 65-year olds; in contrast to 70% and 11%, respectively, caused by GII.4. GII.17 is reported to be an immune-escape variant with a high evolutionary rate (4.4×10^{-2} nucleotide substitutions per site per year; 10-fold higher than GII.4), which could account for its rapid emergence in Asia [132]. It is yet to be seen whether GII.17 will replace GII.4 globally; nevertheless early detection of GII.17, facilitated by genotyping, will allow public health bodies to prepare for potential heightened norovirus activity in coming seasons.

GII.4 outbreaks are likely to represent person-person transmission [133]; conversely its association with person-borne transmission means that only 10% of foodborne outbreaks are associated with GII.4 [134]. GI noroviruses on the other hand are rarely detected in hospital settings; associated with only 2% of hospital infections in Denmark between 2002 and 2010 [130]. Instead GIs are more frequently detected in, and therefore more likely to indicate, foodborne outbreaks [133]. In the USA GI.3, GI.6, GI.7, GII.3, GII.6 and GII.12 are two to seven times more likely to be detected in food-borne rather than person-to-person transmission [135].

Even though only 14% of all norovirus outbreaks globally are foodborne, 37% of outbreaks involving mixed genotypes are considered foodborne [134]; a diversity of genotypes is likely to indicate contamination at source with sewage [136-138].

The detection of mixed genotypes in sewage suggests that there is a far greater diversity of norovirus genotypes circulating in the community than in hospital settings. Community acquired infections (CAIs) which are detected in hospitalised patients show a broader range of genotypes than those which are hospital acquired (HAI), with 17% compared to 6% non-GII.4 infections, respectively [130]. However CAIs which are detected in hospital are not a true reflection of the norovirus genotypes circulating in the community as they signify a presentation bias; they do not represent acute sporadic infections in patients without co-morbidities. Therefore the true distribution of norovirus genotypes in the community is not known.

1.3 MOLECULAR METHODS

1.3.1 The clinical application of molecular epidemiology

To establish transmission events, classical epidemiology relies on linking norovirus cases in time and place; i.e. occurring physically near each other (such as the same ward) and at a similar time (within the incubation period; up to 2-3 days). Molecular epidemiology, on the other hand, compares genomic sequences of the virus, be it partial or full genomes, to determine whether they are similar enough to be linked cases.

A full understanding of transmission dynamics is critical for prevention and control of norovirus transmission, both in food-borne and healthcare settings. Classical epidemiology, which links cases in time and place, is important in identifying outbreaks and, based on circumstantial evidence, may give an indication of the source of transmission. Nevertheless circumstantial evidence may incorrectly link cases or, conversely, not identify transmission between cases that are not traditionally linked in time or place. The implication of these scenarios is potential miss-allocation of infection control resources in the former or unrecognised transmission, potentially leading to a wider outbreak, in the latter.

Knowledge of the viral genome in the context of transmission and outbreaks, termed molecular epidemiology, may provide an evidence base for outbreak investigations, and thus the implementation of interventions. Classical genotyping techniques provide useful information, in particular to identify the genotypes causing infection, however the technique is time-consuming and only provides a snapshot of information; in particular

genotyping is of limited use when a predominant genotype is in circulation, as is the case with norovirus GII.4 globally. Full genome sequences, on the other hand, provide us with greater resolution; in addition to the known regions of interest used in conventional genotyping, such as the RNA polymerase and capsid sequences, full genomes allow us to identify recombination events and to analyse minority variants in the intra-host viral population. The data generated from full genome sequencing has the potential to unequivocally include or exclude cases from an outbreak and, especially in food-borne point source outbreaks, confirm the source of infection.

Norovirus is best known for causing large outbreaks in enclosed settings such as schools, cruise ships, healthcare and military facilities; these are usually caused by a point-source contamination and/or person-to-person transmission. However foodborne outbreaks can sometimes affect several towns, states or countries. In the case of multi-region food-borne outbreaks caused by the same contaminated food-stuff, cases may not be recognized as part of a wider outbreak using classical epidemiology alone.

In instances where foodborne outbreaks are dominated by a single, rare, genotype, genotyping can be used to link cases across geographical regions. For instance a peak in norovirus cases in Denmark over a three month period in 2010/2011 was initially attributed to six independent outbreaks. However genotyping using partial polymerase and capsid sequences across these outbreaks revealed a shared genotype, GI.Pb_GI.6, which was also detected in a batch of imported frozen raspberries; thus linking the outbreaks and implicating the raspberries as the common source [139].

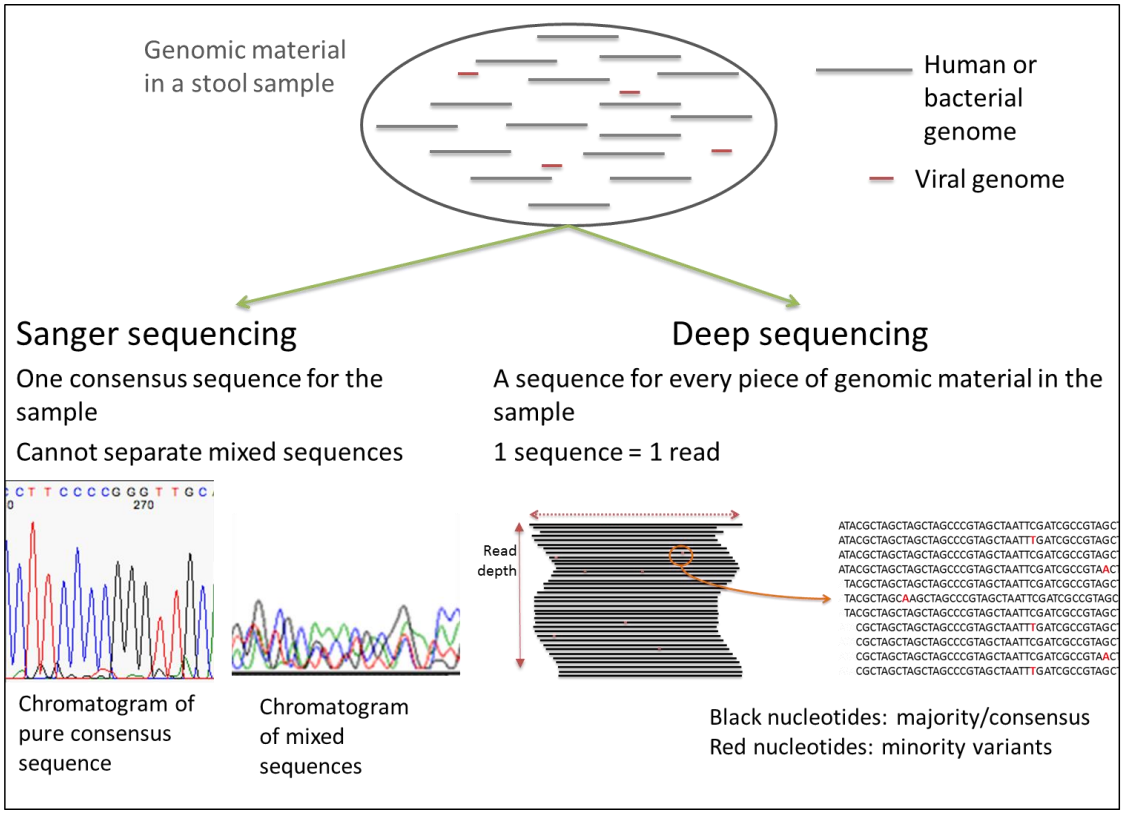
In a healthcare setting, molecular epidemiology allows us to compare viral genome sequences between cases, thus excluding cases from an outbreak or linking previously unrecognised transmission events. The application of molecular epidemiology to elucidate nosocomial transmission was elegantly demonstrated in a Dutch hospital, in which the comparison of capsid P2 sequences from patients infected with norovirus revealed that outpatients, who visit the hospital for short appointments therefore were assumed to have infections from the community, were shown in fact to be linked to outbreaks within the hospital. This consequently allowed targeting of infection control practices in outpatient departments, which were previously not recognised as a risk for nosocomial transmission [140].

1.3.2 Sequencing methods: Capillary and deep sequencing

Capillary (Sanger) sequencing uses chain-terminating ddNTPs followed by capillary electrophoresis to produce a single consensus sequence from a PCR amplicon. Mixed sequences cannot be resolved (**Figure 1.8**).

Conversely deep sequencing, also known as next generation sequencing (NGS), generates a single sequence from each fragment of DNA present in the sequencing library. This means mixed sequences can be resolved. Read depth refers to the number of sequencing reads at each position, be it across a genome or PCR amplicon. Whilst Sanger sequencing effectively has a read depth of one (i.e. the consensus sequence only), deep sequencing can have up to several thousand reads per position. A consensus sequence can be derived from deep sequencing reads by determining the nucleotide that is present in the majority of reads at that position. Conversely a minority variant is a nucleotide that is present in a minority of reads at that position (**Figure 1.8**).

Figure 1.8 Schematic demonstrating the difference between capillary (Sanger) and deep sequencing



1.3.3 Norovirus genotyping

Conventional genotyping for norovirus is based on PCR amplification and capillary sequencing of 330–650 nt fragments of the norovirus polymerase or capsid genes, as illustrated in **Figure 1.2** on page 33. Capsid genotyping is achieved by PCR amplification of the shell domain using genogroup-specific primers. The broad-range nature of these primers, necessary due to sequence variation between genotypes, results in limited sensitivity therefore a nested PCR approach is often required. Following capillary sequencing of the capsid shell domain amplicon, the genotype is determined by reconstructing a phylogeny of the unknown sequence with reference

sequences of known genotypes. The genotype is assigned based on clustering with a known reference with bootstrap support >70 [18]. The same PCR and phylogeny approach is taken for polymerase typing. Both analyses can be done using the web-based Norovirus Genotyping Tool [141], which is a publically available online repository of norovirus sequences, to which one submits the polymerase and/or capsid consensus sequence of an unknown sample, which automatically re-constructs a phylogeny for each and reports the genotype.

Once the capsid genotype is determined, genotype-specific primers are used to amplify the variable P2 region. Universal GI and GII primers to amplify P2 have been reported [138], negating the need for different primer sets to amplify each of the 31 known norovirus GI and GII genotypes; these broad-range primers can successfully amplify the P2 region of several genotypes however with only a 71% success rate. Consequently a separate primer set for each genotype is common. The draw-back of this approach is that multiple primer sets are required (31 for all genotypes), the genotype must be known and, even so, primer mismatches due to sequence diversity within genotypes may still cause the PCR to fail.

Once sequenced, the P2 region can be used in two ways. First, to determine the variant type in GII.4 sequences using the online Norovirus Genotyping tool. Similarly to the polymerase and shell domain sequences, the P2 sequence from a GII.4 norovirus can be compared to known reference sequences in a phylogeny to determine the GII.4 variant type, for example Sydney_2012 or New Orleans_2009.

Second, the sequence identity of the P2 region (approx. 430 nt) from samples of interest can be compared and used in phylogenetic analysis to determine if several norovirus episodes are linked. This can be useful in an immunocompetent population in whom linked episodes are expected to have identical P2 regions [138, 142, 143], but of limited use in immunocompromised populations in whom there is extensive intra-host diversity [52] and therefore P2 regions may not be identical between linked episodes [144]. In the context of an immunocompromised population full genome sequences may provide added resolution and thus be more informative to recognise transmission.

1.3.4 Utility of whole genome sequencing

Traditional polymerase and capsid genotyping by PCR and additional sequencing of the P2 region for outbreak investigations is a labour intensive process requiring several rounds of PCR and sequencing, each requiring genogroup or genotype specific primers, which only yields partial genome sequences at the end. Moreover whilst the P2 domain can identify linked outbreak events with 64–73% specificity (assuming

bootstrap support >70 or <70, respectively), the full capsid sequence can identify linked outbreak events with 100% specificity [145] thus is more informative.

Whole genome sequencing simplifies investigation of norovirus molecular epidemiology by generating all the regions of interest in one step, thus allowing identification of the genotype, variant type, and full capsid sequence; negating the need for sequential PCR and sequencing reactions.

By increasing the sequence data available, whole genome sequencing for molecular epidemiology has the potential to increase our understanding of the transmission dynamics of norovirus. This can be applied to develop evidence-based interventions and target infection control practices; thus minimising transmission and reducing the burden of norovirus disease.

1.3.4.1 Full genomes reveal inter- and intra-genotype recombination

Whole genome sequencing generates a full capsid sequence but, crucially, also the full ORF 1 sequence. This is important because recombination between norovirus genomes occurs with a breakpoint in or near the ORF1/ORF2 overlap region [19, 20] therefore to get a full picture of norovirus epidemiology both ORF1 and ORF2 must be analysed, which is facilitated by whole genome sequencing. Recombination is particularly important in the emergence of new GII.4 variants. It has been proposed that inter- and intra-genotype recombination is likely to be an important force in driving the evolution and emergence of novel GII.4 variants, by affecting the antigenic properties of a variant (via acquisition of a different ORF2/3 sequence) or by altering the balance of replication and mutation rates, thus increasing viral fitness (via acquisition of a different ORF1). All of the GII.4 variants since Farmington Hills_2002 have been influenced by recombination, either as the parent of a new recombinant or as a recombinant itself; the current dominant variant GII.Pe_GII.4 Sydney_2012 is a recombinant of Osaka_2007 (ORF1) and Apeldoorn_2008 (ORF 2/3) [146].

The utility of whole genome sequencing to identify recombinant strains was demonstrated by Wong *et al.* [147] who sequenced 32 stool samples from patients infected with norovirus GII.4 in England during the 2012/2013 winter season, achieving full genome sequences in 23 of the 32 samples. The winter season of 2012/2013 saw the GII.4 variant Sydney_2012 replace the previously circulating variant New Orleans_2009; Wong and colleagues used full genome sequences to identify two stool samples with New Orleans_2009/Sydney_2012 recombinant sequences, suggesting that during co-circulation of the two variants co-infection and recombination had occurred. Wong *et al*. identified the recombination breakpoint in the ORF1/ORF2

overlap region, which confirms the breakpoint previously identified using partial genome sequences by Bull *et al* [19].

Detection of recombination is important for local epidemiology, as well as global. An outbreak investigation that utilises capsid sequences alone will not identify recombinant sequences; patients infected with recombinant virus will be miss-identified as belonging to one transmission cluster, when in fact they are linked to more than one.

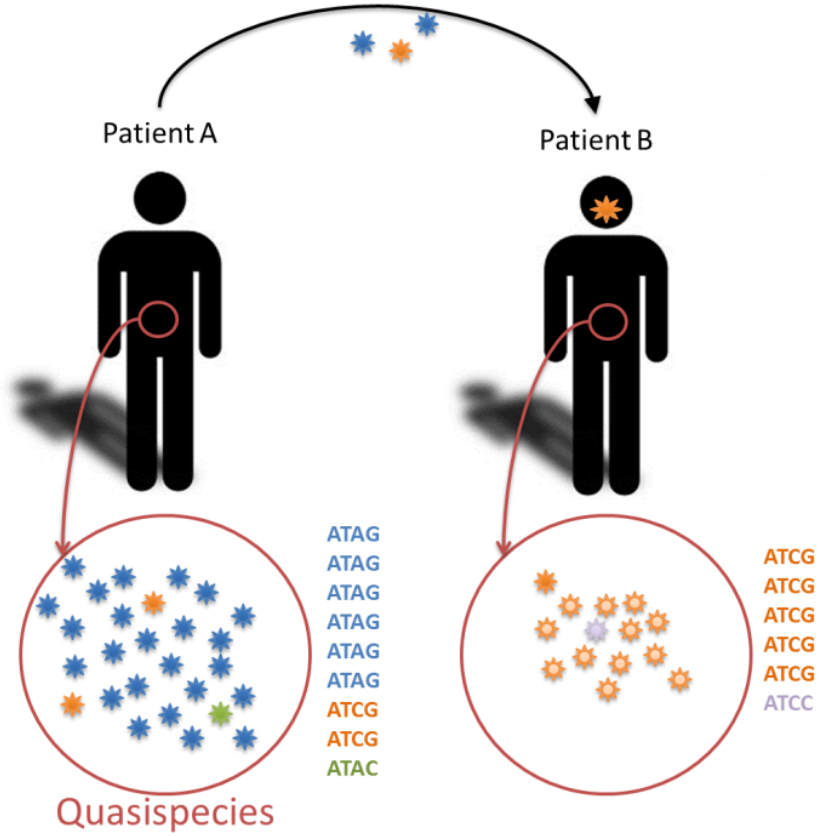
1.3.4.2 Intra-host minority variants suggest direction of transmission

One additional utility of generating full genome sequences using deep sequencing methods is that it allows analysis of minority variants.

RNA viruses lack proof-reading enzymes, resulting in a high mutation frequency of 10^{-3} to 10^{-5} substitution per nucleotide per replication cycle [21, 51]. Intra-host viral populations consequently exist as a heterogeneous population with a consensus sequence and minority variant sequences, known as quasispecies. During a norovirus transmission event between an infected donor and an uninfected recipient, only some of those variants will establish a new infection in the recipient [52]. By comparing the consensus sequence and minority variants between two linked patients, it may be possible to infer the direction of transmission (i.e. which is the donor and which is the recipient) by demonstrating that minority variant sequences in one patient (the donor) are seen as the consensus sequence in the other patient (the recipient), as illustrated in **Figure 1.9**. Full genomes are required for this kind of analysis as the minority variant sites can be spread across the genome [148].

Whilst molecular data can be highly suggestive of transmission, including the direction of transmission [148], it cannot unequivocally prove direct transmission between two individuals since an unidentified intermediary may be involved in the transmission chain. Moreover one should not exclude the possibility that two individuals in whom the molecular evidence is suggestive of transmission may have independently acquired infections with similar, but not directly linked, norovirus strains.

Figure 1.9 Illustration of norovirus minority variant analysis and transmission dynamics. A minority variant from Patient A (orange) is transmitted to Patient B, in whom an infection is established. Deep sequencing of the quasi-species in each patient shows that the nucleotide sequence corresponding to the minority variant in Patient A is seen as the majority (consensus) sequence in Patient B



1.3.4.3 Estimates of norovirus evolutionary rate using full genomes

The largest number of norovirus whole genome sequences generated to date in a single study has allowed estimations of norovirus evolutionary rates [149]. More than 100 whole genome sequences (n=112) from stool samples collected in Vietnam between 2009 and 2011 suggest the evolutionary rate across the whole genome is 5.34 – 6.15 $x10^{-3}$ substitutions per site per year, which supports previous estimates generated using partial genome sequences. Analysis of full genomes showed that ORF 1, encoding a non-structural polyprotein including RNA-dependent RNA polymerase, exhibited a lower evolutionary rate than the major capsid protein encoding ORF 2. Moreover for the first time the evolutionary rate was calculated for ORF3, which is a minor capsid protein with poorly understood function; it was demonstrated that ORF 3 was the region with the highest evolutionary rate across the genome (7.38–8.99 $x10^{-3}$ substitutions per site per year), although this difference was not statistically significant due to overlapping 95% confidence intervals. The authors propose ORF 3 could

therefore be a possible site of virus-host interaction under selective pressure. Without full genome sequences this would remain unrecognised. Additionally, whole genome sequencing of several genotypes in this study suggests similar evolutionary rates in GII.3 and GII.4 sequences. Infections with GII.4 are far more prevalent than GII.3; it has previously been suggested that this could be due to a higher evolutionary rate in GII.4 viruses [150]. A similar evolutionary rate in the two genotypes suggests the difference in prevalence could instead be a sampling bias caused by infections with GII.4 genotypes presenting more frequently to healthcare facilities, although there is currently not any genotyping data available from community cohorts to corroborate this.

1.4 AIMS AND OBJECTIVES

The primary aim of this thesis is to assess the utility of norovirus next generation sequencing in clinical practice, in the context of an immunocompromised paediatric population. This will be achieved via the following objectives.

Objective 1: Determine the burden of norovirus infection in the study population, in the context of other gastrointestinal viral infections (Chapter 3)

Objective 2: Evaluate the performance of a novel method for sequencing norovirus full genomes, namely target enrichment using SureSelect [Agilent] (Chapter 4)

Objective 3: Use norovirus full genome sequencing to investigate the molecular epidemiology of norovirus in a nosocomial paediatric population (Chapter 5)

Objective 4: Use deep sequencing to evaluate the efficacy of a broad spectrum antiviral, Ribavirin, in the treatment of chronic norovirus infections (Chapter 6)

CHAPTER 2

GENERAL MATERIALS AND METHODS

2.1 GREAT ORMOND STREET HOSPITAL FOR CHILDREN (STUDY POPULATION)

Great Ormond Street Hospital (GOSH) is a tertiary referral paediatric hospital in central London, UK, that receives over 268,000 patient visits per year. GOSH has 350 beds, 60% of which are in isolation. Most of the children seen at the hospital are referred from other hospitals throughout the UK and overseas. There is no accident and emergency department therefore acute gastroenteritis is not the primary reason for admission.

The hospital has more than 50 clinical specialties with a focus on highly specialist services, including those that involve patients with primary and secondary immunodeficiencies.

GOSH is one of the largest centres for bone marrow transplant (BMT) for severe immunodeficiency worldwide and the largest paediatric/adolescent oncology centre in Europe, treating approximately 450 haematology/oncology patients each year.

Each year GOSH performs over 100 hematopoietic stem cell transplants (HSCT), primarily for patients from oncology, haematology and immunology clinical specialties. In addition GOSH is a renal, cardiac and lung transplant centre.

2.2 ROUTINE DIAGNOSTIC TESTING FOR GASTROINTESTINAL VIRUSES

As part of the infection control screening policy at GOSH, all children are tested for gastrointestinal viruses on admission for inpatient stay, regardless of whether they are symptomatic or asymptomatic. Stool samples are tested by real-time PCR for the presence of norovirus, adenovirus, rotavirus sapovirus and astrovirus. Patients with a positive stool virus are followed up weekly until they become negative, however, patients with underlying immunodeficiency are tested weekly and thereafter at outpatient appointments irrespective of whether they have previously been positive for a stool virus. Any child who develops gastrointestinal symptoms during their inpatient stay or in outpatients is also tested. Before 2010 stools were tested for the presence of viruses by electron microscopy. Since 2010 stools are tested by reverse transcription real-time PCR.

Norovirus infections detected less than 48 hours after admission to hospital are considered positive on admission (POA); detection more than two days after admission is considered a hospital acquired infection (HAI). Since many of the patients at GOSH

have complex medical histories many of them have previously been admitted to local hospitals or had several outpatient visits prior to admission at GOSH; therefore earlier acquisition of infection in another healthcare facility cannot be excluded.

Norovirus infections detected for longer than one month are considered chronic infections; less than one month are considered acute.

2.3 INFECTION CONTROL POLICY AT GOSH

At GOSH there is an Infection Prevention and Control (IPC) team who are responsible for coordinating the surveillance, investigation and prevention of infection in patients, staff and visitors. This team is led by the director of infection prevention and control (DIPC), typically a clinician, and additionally includes three IPC nurses and one IPC Clinical Scientist.

On a daily basis the IPC team are informed by the diagnostic Virology laboratory of all new episodes of norovirus detected by real-time PCR. The IPC team reviews these results daily for any evidence of cross infection. In addition the IPC team are informed by ward staff of any patients with diarrhoea and/or vomiting. An outbreak of gastroenteritis is suspected of occurring when an epidemiologically linked group of patients, staff or visitors develop symptoms associated with gastroenteritis or an indistinguishable organism, such as norovirus, is detected from them.

Once an outbreak is suspected an incident report is completed and an outbreak control group may be formed to coordinate actions. Any symptomatic staff are excluded from duty, leave the hospital and do not return until they are asymptomatic for 48 hours. Patients who are either symptomatic or have norovirus detected in stool are isolated if they are not already and if there is a room available. The decision to close units or wards or to restrict admissions is based on the risk to the patient population and the number of isolation facilities available.

2.4 RNA EXTRACTION FROM STOOL

A pea-sized amount of stool was mixed with 1 ml stool lysis buffer (Qiagen Buffer ASL), vortexed, and centrifuged for 5 minutes at 13,000 rpm to pellet cellular debris. The clarified stool suspension was used for RNA extraction.

RNA was extracted from 200 µl of the clarified stool suspension using the Qiagen EZ1 virus mini kit or Qiasymphony DSP Virus/Pathogen kit with a 90 µl elution volume. All purified RNA was stored at −80°C prior to cDNA synthesis.

2.5 REAL-TIME PCR

Real-time PCR reactions for norovirus, rotavirus, astrovirus and sapovirus consisted of 10 µl RNA, primers and probes as per **Table 2.1**, 12.5 µl Qiagen Quantifast RT PCR mastermix, 0.25 µl RT enzyme and water to make a total volume of 25 µl. One-Step PCR cycling consisted of 50 ^{o}C for 20 minutes (cDNA synthesis) and 95 ^{o}C for 5 minutes (RT enzyme inactivation and activation of hot-start *Taq* polymerase), followed by 45 cycles of 95 ^{o}C for 15 seconds (denaturation) and 60 ^{o}C for 30 seconds (primer annealing and template extension). Real-time PCR reactions for adenovirus consisted of 10 µl DNA, primers and probe as per **Table 2.1**, 12.5 µl Qiagen Quantifast PCR mastermix and water to make a total volume of 25 µl. Cycling consisted of 95 ^{o}C for 5 minutes followed by 45 cycles of 95 ^{o}C for 30 seconds and 60^{o}C for 30 seconds. All PCR targets have a single copy in the genome. A positive and negative template control was included in every PCR run to ensure the reliability of results.

The real-time PCR assay to detect norovirus GI and GII was developed by me shortly before the commencement of the work contributing to this thesis. The assay validation methods are described in section 2.5.1 on page 64 and a summary of the results are shown in section 3.3.1 on page 87.

The real-time PCR assays to detect rotavirus, adenovirus, sapovirus and astrovirus were previously developed by the GOSH Virology Department. These assays, and the norovirus assay described above, are routinely used by the GOSH virology department for the diagnosis of gastrointestinal infections. The ongoing performance of each real-time PCR assay is monitored by the GOSH Virology diagnostic laboratory through regular Internal Quality Assurance (IQA), annual External Quality Assurance (EQA) and daily monitoring of positive control Ct values (with strict acceptability criteria) to ensure they remain fit for purpose.

IQA consists of blindly re-submitting specimens for diagnosis of infection to ensure the obtained results are reproducible. EQA consists of a commercially distributed panel of specimens containing target pathogens to ensure each pathogen is correctly identified, with no cross-reactivity, false positive or false negative results.

The real-time PCR assays for detection of gastrointestinal viruses described here meet the quality and competence requirements for accreditation to the internationally recognised standard ISO:15189 (Medical Laboratory Accreditation), as assessed by the United Kingdom Accreditation Service (UKAS).

Table 2.1 Primer and probe sequences used for PCR assays

Assay name [Reference]	Primer/probe name [Reference]	Oligonucleotide sequence	Final conc. (μM)	Target and amplicon size	Lower limit of detection* (and assay validation**)
Norovirus real-time RT-PCR	COG-1F [151]	CGYTGGATGCGNTTYCATGA	0.8	ORF1/ORF2 junction, 86–89 nt	GI: 5 copies/μl
	COG-1R [151]	CTTAGACGCCATCATCATTYAC	0.8		GII: 1 copy/μl
	RING-1a [151]	FAM-AGATYGCGATCYCCTGTCCA-BHQ1	0.2		(this study)
	RING-1b [151]	FAM-AGATCGCGGTCTCCTGTCCA-BHQ1	0.1		
	QNIF2 [152]	ATGTTCAGRTGGATGAGRTTCTCWGA	0.8		
	COG-2R [151]	TCGACGCCATCTTCATTCACA	0.8		
	QNIFS [152]	Cy5-AGCACGTGGGAGGGCGATCG-BHQ2	0.2		
Rotavirus real-time RT-PCR [153]	Rota-F	ACCATCTWCACRTRACCCTCTATGAG	1	NSP3, 87 nt	5 copies/μl
	Rota-R	GGTCACATAACGCCCCTATAGC	1		(GOSH Virology
	Rota-Pr	FAM-AGTTAAAAGCTAACACTGTCAAA-MGB	0.5		department)
Sapovirus real-time RT-PCR [154]	SaV-F1	GAYCASGCTCTCGCYACCTAC	0.4	ORF1/ORF2 junction, 103 nt	3 copies/μl
	SaV-F2	TTGGCCCTCGCCACCTAC	0.4		(GOSH Virology
	SaV-F3	TTTGAACAAGCTGTGGCATGCTAC	0.4		department)
	SaV-R	CCCTCCATYTCAAACACTA	0.4		
	SaV-Pr1	FAM-CCRCCTATRAACCA-MGB	0.2		
	SaV-Pr2	FAM-TGCCACCAATGTACCA-MGB	0.2		

Assay name [Reference]	Primer/probe name [Reference]	Oligonucleotide sequence	Final conc. (µM)	Target and amplicon size	Lower limit of detection* (and assay validation**)
Astrovirus real-time RT-PCR [155]	Astro-F	TCAACGTGTCCGTAAMATTGTCA	0.8	Capsid, 66 nt	1 copy/µl (GOSH Virology department)
	Astro-R	GCWGGTTTTGGTCCTGTGA	0.8		
	Astro-Pr	VIC-CAACTCAGGAAACARG-MGB	0.5		
Adenovirus real-time PCR	AdenoUCL-F	GCCACSGTGGGGTTTCTAAACTT	0.6	Hexon gene, 132 nt	0.2 copies/µl (GOSH Virology department)
	AdenoUCL-R	GCCCCAGTGGKCTTACATGCACATC	0.6		
	AdenoUCL-Pr	FAM-TGCACCAGACCCGGRCTCAGGTACTCCGA-BHQ1	0.4		
Norovirus GI first round Genotyping PCR [129, 143]	GIFF-1	ATHGAACGYCAAATYTTCTGGAC	0.5	Capsid shell domain, 597 nt	(This study)
	GIFF-2	ATHGAAAGACAAATCTACTGGAC	0.5		
	GIFF-3	ATHGARAGRCARCTNTGGTGGAC	0.5		
	GISKR	CCAACCCARCCATTRTACA	0.5		
Norovirus GI second round Genotyping PCR (semi-nested) [129, 143]	GIFFN	GGAGATCGCAATCTCCTGCCC	0.5	Capsid shell domain, 364 nt	(This study)
	GISKR	as above			
Norovirus GII first round Genotyping PCR [129, 143]	GIIFB-1	GGHCCMBMDTTYTACAGCAA	0.5	Capsid shell domain, 468 nt	(This study)
	GIIFB-2	GGHCCMBMDTTYTACAAGAA	0.5		
	GIIFB-3	GGHCCMBMDTTYTACARNAA	0.5		
	GIISKR	CCRCCNGCATRHCCRTTRTACAT	0.5		

Assay name [Reference]	Primer/probe name [Reference]	Oligonucleotide sequence	Final conc. (μM)	Target and amplicon size	Lower limit of detection* (and assay validation**)
Norovirus GII second round Genotyping PCR (semi-nested) [129, 143]	GIIFBN	TGGGAGGGCGATCGCAATCT	0.5	Capsid shell domain, 343 nt	(This study)
	GIISKR	as above			
NS7 PCR	g2-4F4404 [156]	TCCTCAGAACCACATTTGGCTCAGGTAGTCGCAGAAG	0.2–0.5	RNA polymerase, 739 nt	(This study)
	g2-4R5115 [156]	ACGAGGTTGGCTGTGGACCCATCAGATG	0.2–0.5		
VP1 PCR	g2-4F5686 [156]	ATGATGTCTTCACAGTCTCTTGTCG	0.2–0.5	Major capsid protein, 857 nt	(This study)
	g2-4R6515 [156]	ACCCTACCTGTGTCTGGATTCACAAATC	0.2–0.5		
Primers for reverse transcription	UNP1-163 †	CAGCCGTGTGGACTCCAA	0.03	n/a	(This study)
	UNP1-253 †	GGGTGGCACATATGACAGTGTT	0.03		
	UNP1-82 [149]	GACCTCTGGGACGAGGTTG	0.03		
	UNP1-135 [149]	CTCCACCAGGGGCTTGTAC	0.03		
	UNP1-44 [149]	GCACGGTTGAGACTGTGC	0.03		
	UNP1-100 [149]	GCCAGTCCAGGAGTCCAA	0.03		

n/a, not applicable

* For real-time PCR assays only

** assay validation refers to assay design using published primer sequences, optimisation and validation of performance characterisation

† unpublished; personal communication Paul Kellam (Sanger Institute)

2.5.1 Validation of performance characteristics for norovirus real-time PCR (RT-qPCR)

The *in silico* specificity of the norovirus GI and GII primers and probes was verified by submitting each of the primer and probe sequences to NCBI nucleotide BLAST (https://www.ncbi.nlm.nih.gov/BLAST) to compare the primer and probe sequences to all sequences in the Genbank database.

To determine specificity of the multiplex norovirus real-time PCR assay, targeting GI and GII, the assay was tested with stool samples known to be positive for rotavirus, adenovirus, astrovirus and sapovirus (one stool per pathogen).

To determine the limit of detection, the RT-qPCR assay was tested on a ten-fold dilution series of plasmid containing the PCR target sequence. A separate plasmid was constructed for the GI and GII targets. Each plasmid was tested from 1–1x10^6 copies/µl. Plasmids were prepared by cloning using the TOPO TA Cloning kit with chemically competent *Escherichia coli* cells [Invitrogen] and quantified by Nanodrop Spectrophotometer.

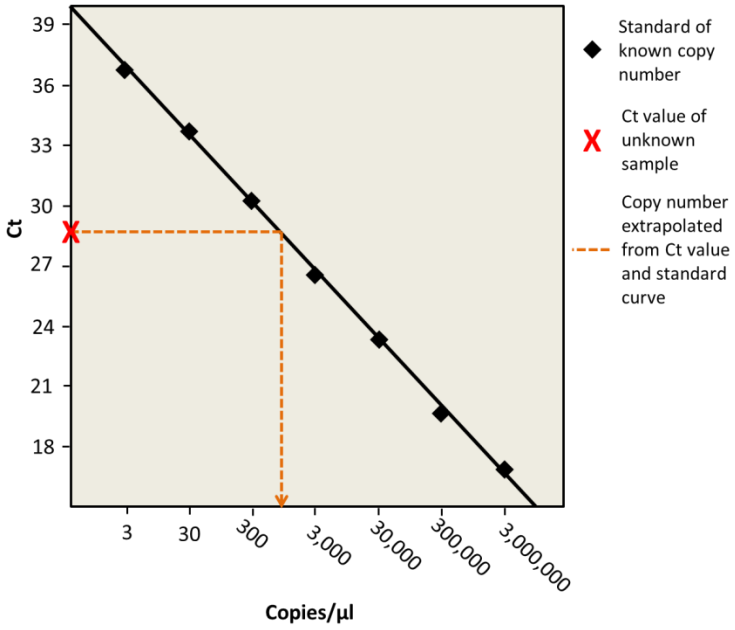
To assess the sensitivity of the real-time assay, five GII, four GI and one GIV positive stool samples, supplied by Quality Control for Molecular Diagnostics (QCMD), were tested using the norovirus real-time assay. QCMD is an external quality assessment (EQA) programme used to monitor and improve the performance of diagnostic assays. The samples in each QCMD panel are designed to resemble clinically significant specimens.

To test the reproducibility of the norovirus RT-qPCR and the relationship between Ct values and viral load, a ten-fold dilution series of norovirus GII plasmid from 3 to 3x10^6 copies/µl was tested on four separate days.

2.5.2 Viral load

A 10-fold dilution series of the plasmid DNA containing the PCR target sequence, from 3 to 3x10^6 copies/µl, was run in parallel to extracted stool samples to construct a standard curve. The viral load in copies/µl is determined by comparing the Ct value of the tested sample to the standard curve, as illustrated in **Figure 2.1**.

Figure 2.1 Viral load calculation using real-time PCR



2.6 PREPARATION OF cDNA

2.6.1 For gel-based capsid genotyping PCR (Chapter 3)

cDNA was prepared using random primers and SuperScript III reverse transcriptase enzyme (Invitrogen). Briefly, reactions consisted of 18.5 ng/µl random primers, 0.77 µM each dNTP, 5 µl RNA and water to make volume up to 13 µl. Reactions were heated to 65 ^{o}C for 5 minutes then cooled on ice. To each reaction 4 µl first-strand buffer and 1 µl each of 0.1M DTT, SuperScript III (Invitrogen) and RNase OUT (Invitrogen) was added. Reactions were incubated at 25 ^{o}C for 5 minutes, 50 ^{o}C for 1 hour and 70 ^{o}C for 15 minutes. 0.6 µl RNase H (New England Biolabs) was added to each reaction prior to a final incubation of 37 ^{o}C and 65 ^{o}C for 20 minutes each to remove residual RNA.

2.6.2 For SureSelect target enrichment and deep sequencing (Chapter 4)

RNA extracts (90 µl) were concentrated to 11 µl using a vacuum centrifuge at 65^{o}C prior to first strand cDNA synthesis. First strand cDNA was synthesised using random primers and SuperScript III (SS III, Life Technologies) as per manufacturer's instructions. Briefly, 1 µl of 10mM (each) dNTP mix and 1 µl of 3 µg/ml random primers were incubated with 11 µl RNA for five minutes at 65 °C to anneal primers to RNA template, followed by incubation on ice for 1 minute. RNA–primer templates were mixed with 4 µl 5x first strand buffer, 1 µl 0.1M DTT, 1 µl RNase OUT and 1 µl SS III at 25 °C for 5 minutes followed by cDNA synthesis at 50 °C for 1 hour and enzyme inactivation at 70 °C for 15 minutes. Second strand cDNA was synthesised using

Second Strand cDNA Synthesis kit (NEB) as per manufacturer's instruction. Briefly, 20 µl first strand cDNA was incubated with 48 µl water, 8 µl 10x 2nd strand buffer and 4 µl 2nd strand enzyme mix at 16 °C for 2.5 hours. Double stranded cDNA was purified and concentrated with Genomic DNA Clean and Concentrator (Zymo Research), as per manufacturer's instructions, with a 30 µl elution volume and quantified with Qubit dsDNA high sensitivity (HS) kit (Invitrogen).

2.6.3 For gel-based NS7 and VP1 PCR (Chapter 6)

cDNA was prepared using universal norovirus GII.4 primers (UNP1, **Table 2.1**) and SuperScript III reverse transcriptase enzyme (Invitrogen). Briefly, reactions consisted of 0.03 µM each primer, 0.77 µM each dNTP, 20 µl RNA and water to make volume up to 65 µl. Reactions were heated to 65 ^{o}C for 5 minutes then cooled on ice. To each reaction 20 µl first-strand buffer and 5 µl each of 0.1M DTT, SuperScript III (Invitrogen) and RNase OUT (Invitrogen) was added. Reactions were incubated at 25 ^{o}C for 5 minutes, 50 ^{o}C for 1 hour and 70 ^{o}C for 15 minutes. 3 µl RNase H (New England Biolabs) was added to each reaction prior to a final incubation of 37 ^{o}C and 65 ^{o}C for 20 minutes each.

2.7 GEL-BASED PCR

2.7.1 PCR for capsid genotyping (Chapter 3)

A 468–597 nt fragment of the capsid shell domain was amplified by PCR using primers detailed in **Table 2.1**. Briefly, reactions consisted of 5 µl cDNA, 0.5 µM each primer, 2.5 µl Accuprime Buffer I (Invitrogen), 0.5 µl Accuprime Taq polymerase and water to make volume to 20 µl. Denaturation at 94 °C for 2 minutes was followed by 40 cycles of 94 °C for 30 seconds, primer annealing at 42°C for 60 seconds and template extension at 68°C for 90 seconds. Samples that were negative by first round PCR were amplified by second round semi-nested PCR using primers as listed in **Table 2.1** and the same cycling conditions as the first round.

Amplified products were visualised by gel electrophoresis on a 2% agarose gel and capillary sequenced to generate a consensus sequence.

2.7.2 PCR for NS7 and VP1 for cloning and deep sequencing comparison (Chapter 6)

NS7 and VP1 PCR reactions consisted of 0.2 µM each primer, 0.25 µl GoTaq DNA Polymerase (Promega), 10 µl Green GoTaq Reaction Buffer, 2.5 mM $MgCl_2$, 0.2mM each dNTP and water to make volume to 45 µl. Cycling conditions consisted of initial

denaturation at 95 °C for 2 minutes followed by 30 cycles of 95 °C for 30 seconds, 50 °C for 30 seconds and 72 °C for 1 minute and a final extension at 72 °C for 5 minutes. PCR products were visualised on a 2% agarose gel.

2.7.3 PCR for NS7 and VP1 mutation frequency in pre- and post-ribavirin samples (Chapter 6)

NS7 and VP1 PCR conditions were optimised for sensitivity. The final PCR reactions consisted of 0.5 µM each primer, 0.4 µl Accuprime Pfx DNA Polymerase (Invitrogen), 5 µl Accuprime Pfx reaction mix and water to make volume to 40 µl. Cycling conditions consisted of initial denaturation at 95 °C for two minutes followed by 35 cycles of 95 °C for 15 seconds, 52 °C for 30 seconds and 68 °C for 1 minute. PCR products were visualised on a 2% agarose gel.

2.8 GEL ELECTROPHORESIS

For a 2% gel, 1.2 g agarose was dissolved in 60 ml 1X TBE buffer. Once cooled, 7.5 µl ethidium bromide or Gel Red solution (Biotium) was added before casting.

5 µl of PCR product was mixed with 1 µl loading dye (Invitrogen) prior to loading. Every gel was run with 5 µl of Hyperladder IV (Invitrogen) as a size marker. Gels were run at 100 V for 20–30 minutes prior to visualising under UV light.

2.9 CAPILLARY (SANGER) SEQUENCING

Prior to capillary sequencing, PCR products were purified using Illustra MicroSpin columns (GE Healthcare) to remove unincorporated PCR primers and primer-dimer complexes smaller than 200 bp. The MicroSpin columns contain a resin of differing pore sizes; molecules larger than the largest pore size are excluded from the gel matrix and thus are eluted. Smaller molecules are retained by the gel matrix so are excluded from the purified DNA.

All PCR amplicons were sequenced in the forward and reverse direction. Briefly, sequencing reactions consisted of 3 µl purified PCR product, 1.5 µl molecular grade water, 2.5 µl of 2 µM primer (forward or reverse), 1 µl sequencing buffer and 2 µl BigDye 3.1 reaction mix (Life Technologies). Cycling conditions consisted of initial denaturation at 96 °C for 1 minute followed by 25 cycles of 96 °C for 10 seconds, 50 °C for 5 seconds and 60 °C for 4 minutes.

Sequencing products were purified by ethanol precipitation to remove unincorporated primers. The entire sequencing product (10 µl) was mixed with 5 µl 125 mM EDTA and

60 µl 96% ethanol and incubated at room temperature for 15 minutes to precipitate PCR amplicons. Following a 30 minute centrifugation at 2204 g, the pellets were washed with 60 µl 70% ethanol and re-pelleted with a 5 minute centrifugation at 2204 g. Following an inverted pulse centrifugation at 180 g to remove residual ethanol and dry the pellet, pellets were re-suspended in 5 µl HiDi formamide. Sequencing products were denatured for 5 minutes at 95 °C prior to loading onto an ABI 3130 genetic analyser for capillary sequencing.

Sequencing chromatograms were visualised and assembled using Seqman Pro v12 (DNA Star).

2.10 NOROVIRUS GENOTYPING

Consensus sequences (either full genome or partial RNA polymerase/capsid sequences) were submitted to the web-based norovirus typing tool (www.rivm.nl/mpf/norovirus/typingtool) [141], which assigns a genotype based on phylogeny (as described on page 51 in Chapter 1).

2.11 WHOLE GENOME SEQUENCING BY SURESELECT TARGET ENRICHMENT

2.11.1 RNA bait design

Overlapping 120-mer RNA baits complementary to and spanning the length of 622 norovirus partial or complete genomes from Genbank were designed using an in-house PERL script. Briefly, a 120 nucleotide sliding window is scanned along each reference genome at intervals of 10 nucleotides. If the 120-mer is sufficiently different to other 120-mer sequences in the bait set (as assessed by BLAT [157]), it is retained in the bait set; otherwise that 120-mer is discarded. In this way, the bait set spans the diversity in all of the included reference genomes. The reference genomes included polymerase genotypes GI.P1, GI.P2, GI.P3, GI.P4, GI.P6, GI.P8, GI.Pb, GI.Pc, GI.Pd, GI.Pf, GII.P1, GII.P2, GII.P3, GII.P4, GII.P5, GII.P6, GII.P7, GII.P8, GII.P11, GII.P12, GII.P15, GII.P16, GII.P17, GII.P18, GII.P21, GII.P22, GII.Pc, GII.Pe, GII.Pg, GII.Pp, GIII, GIV, GV and GVI and capsid genotypes GI.1, GI.2, GI.3, GI.4, GI.5, GI.6, GI.8, GII.2, GII.3, GII.4, GII.5, GII.6, GII.7, GII.8, GII.10, GII.11, GII.12, GII.13, GII.14, GII.15, GII.16, GII.17, GII.18, GII.21, GII.22, GIII, GIV, GV and GVI. The GII.4 reference genomes included all major GII.4 variants: CHDC1970s, Bristol_1993, Camberwell_1994, US95/96, Farmington Hills_2002, Lanzhou_2002, Asia_2003, Hunter_2004, Yerseke_2006a, Den Haag_2006b, Osaka_2007, Apeldoorn_2007, New

Orleans_2009 and Sydney_2012. The custom designed norovirus bait library was uploaded to Agilent SureDesign and synthesised by Agilent Biotechnologies.

2.11.2 Library preparation, hybridisation and enrichment

Norovirus cDNA samples were quantified and carrier G147 Human Genomic DNA: male (Promega) was added if necessary to obtain a total of 200ng.

All DNA samples were mechanically sheared for 150 seconds using a Covaris E210 focused-ultrasonicator (duty cycle 5%, PIP 175 and 200 cycles per burst) to yield a fragment size of approximately 270 bp. End-repair, non-templated addition of 3' –A adapter ligation, hybridisation, enrichment PCR and all post-reaction clean-up steps were performed according to the SureSelect Illumina Paired-End Sequencing Library XT protocol. All recommended quality steps were performed between steps.

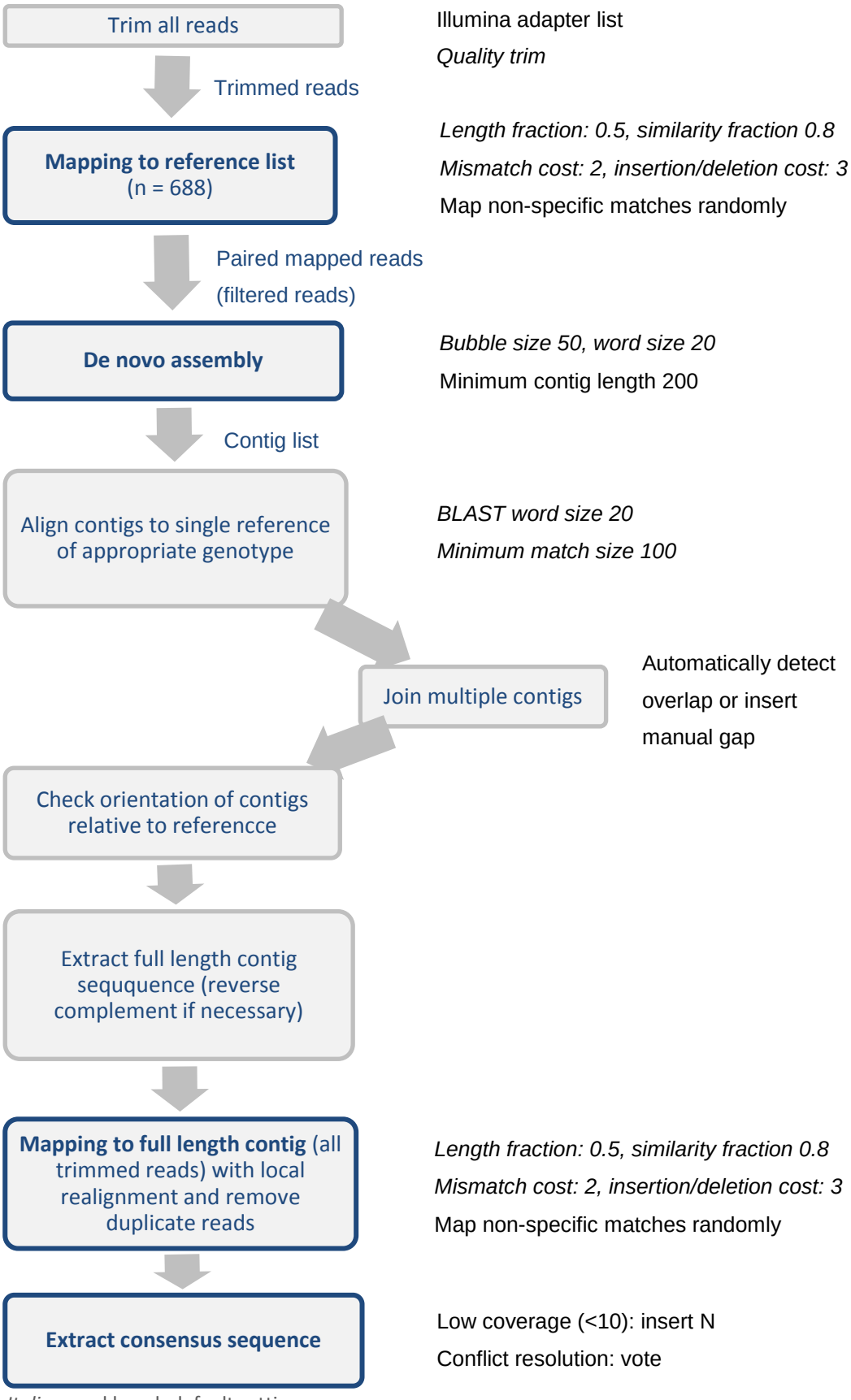
2.11.3 Illumina sequencing of SureSelect enriched libraries

Samples were multiplexed with 48 samples per run. Paired end sequencing was done on an Illumina MiSeq sequencing platform with the 500 cycle v2 Reagent Kit. Base calling and sample demultiplexing were generated as standard on the MiSeq producing paired FASTQ files for each sample.

2.12 WHOLE GENOME ASSEMBLY OF SURESELECT DEEP SEQUENCING DATA

All assemblies were done in CLC genomics workbench v8, as summarised in **Figure 2.2**. All reads were quality trimmed and adapter sequences removed. Trimmed reads were mapped to a curated reference list consisting of all norovirus complete genome and complete gene sequences in Genbank as of 14/07/2015 (n = 688). All paired reads mapping to the reference list (filtered reads) were taken forward to *de novo* assembly using workbench default parameters and a minimum contig length of 200 nucleotides. Contigs generated from the *de novo* assembly were aligned to a single Genbank reference sequence of the relevant genotype to check the orientation of the contig and, where multiple contig sequences were generated, the position of each contig relevant to the reference. Multiple contig sequences were joined based on overlapping nucleotide sequences or with a manually inserted gap. All trimmed reads (pre-filtering) were mapped to the full length contig sequence generated from the *de novo* assembly to generate a final consensus sequence. Areas of low coverage (<10) were assigned the ambiguity symbol N.

Figure 2.2 Schematic of norovirus *de novo* full genome assembly pipeline



Italics, workbench default settings

2.13 IDENTIFICATION OF OPEN READING FRAMES

To verify correct genome assembly, open reading frames (ORF 1, ORF 2 and ORF 3) were identified for all genome sequences using the Find Open Reading Frames tool in CLC Genomics Workbench.

2.14 PHYLOGENETIC ANALYSIS

A phylogenetic tree infers the evolution of a set of taxa, in this instance between norovirus strains from infected patients. Internal nodes represent a common ancestor between two sequences; sequences with a shared common ancestor have a higher degree of relatedness than those that do not have a shared common ancestor. The branch length between nodes represents the genetic distance between two sequences, for instance the average number of nucleotide substitutions per site. The robustness of a tree, i.e. how reliable it is, is estimated using bootstrap values. A bootstrap value indicates the frequency of observing the same topology if the tree were to be reconstructed multiple times; for example if the tree were to be re-constructed 100 times, how many of the resulting trees would have the same topology. Robust branches have high bootstrap values, whilst volatile branches have low values; a bootstrap value of 50 suggests the observed topology will only be replicated in 50% of trees.

All phylogenetic analysis was carried out in CLC Genomics Workbench (v 9.0). Consensus sequences were aligned using the very accurate (slow) progressive alignment tool. Maximum Likelihood phylogenies were reconstructed using the general time reversible (GTR) nucleotide substitution model, including rate variation (+G), tree topology estimation (+T) and 200 bootstrap replicates. The GTR+G+T model was determined to be the best model for the data set by the Model Testing tool in CLC Genomics Workbench. The Model Testing tool applies statistical analysis to trees reconstructed using five different substitution models (see **Table 2.2** for model details) to determine the most suitable for the data set.

Details of all reconstructed phylogenies are summarised in **Table 2.3**. All nodes with bootstrap support <70 were collapsed.

Table 2.2 Summary of nucleotide substitution models tested by the CLC Genomics Workbench Model Testing tool, to determine the most suitable model for Maximum Likelihood phylogeny for a given data set

Model name	Model assumptions
Jukes Cantor (JC)	Equal rates of nucleotide substitution Equal rates of base frequencies
Felsenstein 81	Unequal base frequencies Equal rates of nucleotide substitution
Kimura 80	Transition/transversion substitution bias Equal rates of base frequencies
HKY	Transition/transversion bias Unequal base frequencies
GTR	Transition/transversion bias Unequal base frequencies Unequal rate for each pair of substitutions
+G (applicable to HKY and GTR only)	Site rate variation – i.e. variation is not equal across all sites in the genome
+T	Tree topology estimation to find the tree topology and branch lengths that best describe the sequences' phylogenetic relationships

Table 2.3 Summary of phylogenetic trees reconstructed

Tree details	Number of sequences in tree*	Phylogeny method	Reference sequence	Out-group sequence	Figure No.
GII.4 P2† † (GOSH)	71	Maximum Likelihood	GII.P4_GII.4 New Orleans_2009 JN595867 (USA 2010) GII.Pe_GII.4 Sydney_2012 JX459908 (Australia 2012)	GII.P7_GII.7 KJ196295 (Japan 2010)	Figure 5.20
GI.1 full genome	1	None†	N/A	N/A	N/A
GI.2 full genome	3	Maximum Likelihood	GI.P2_GI.2 KF306212 (China 2013)	GI.P1_GI.1 KF039736 (USA 2010)	Figure 5.11
GI.3 full genome	15	Maximum Likelihood	GI.P3_GI.3 KJ196292 (Japan 2007)	GI.P1_GI.1 KF039736 (USA 2010)	Figure 5.12
GII.1 full genome	1	None†	N/A	N/A	N/A
GII.2 full genome	10	Maximum Likelihood	GII.P16_GII.2 KJ407074 (USA 2011)	GII.P17_GII.17 KT253245 (China 2015)	Figure 5.13
GII.3 full genome**	61	Maximum Likelihood	GII.P12_GII.3 GU980585 (Korea 2006)	GII.P7_GII.7 KJ196295 (Japan 2010)	Figure 5.15
GII.4* full genome**	71	Maximum Likelihood	GII.P4_GII.4 New Orleans_2009 JN595867 (USA 2010) GII.Pe_GII.4 Sydney_2012 JX459908 (Australia 2012) GII.P16_GII.4 Sydney_2012 LC175468 (Japan 2016)	GII.P7_GII.7 KJ196295 (Japan 2010)	Figure 5.16

Tree details	Number of sequences in tree*	Phylogeny method	Reference sequence	Out-group sequence	Figure No.
GII.6 full genome	12	Maximum Likelihood	GII.P7_GII.6 LN854568 (Netherlands 2014)	GII.P17_GII.17 KT253245 (China 2015)	Figure 5.14
GII.7 full genome	3	Maximum Likelihood	GII.P7_GII.7 KJ196295 (Japan 2010)	GII.P12_GII.3 GU980585 (Korea 2006)	Figure 5.18
GII.13 full genome	1	None†	N/A	N/A	N/A
GII.17 full genome	2	Maximum Likelihood	GII.P17_GII.17 KT380915 (Shanghai 2014) GII.P17_GII.17 KP998539 (Hong Kong 2015) GII.P17_GII.17 LC037415 (Kawasaki 2015) GII.P17_GII.17 KT253245 (China 2015) GII.P16_GII.17 KJ196286 (Japan 2002)	GII.P7_GII.6 LN854568 (Netherlands 2014)	Figure 5.17
GII.4 P2† † (NNUH)	60	Maximum Likelihood	GII.P4_GII.4 New Orleans_2009 JN595867 (USA 2010) GII.Pe_GII.4 Sydney_2012 JX459908 (Australia 2012)	GII.P12_GII.3 GU980585 (Korea 2006)	Figure 5.25
GII.4 NNUH full genome	60	Maximum Likelihood	GII.P4_GII.4 New Orleans_2009 JN595867 (USA 2010) GII.Pe_GII.4 Sydney_2012 JX459908 (Australia 2012)	GII.P12_GII.3 GU980585 (Korea 2006)	Figure 5.24
GII.Pe ORF1	40 GII.Pe_II.4 4 GII.Pe_II.2	Maximum Likelihood	GII.Pe_GII.4 Sydney_2012 JX459908 (Australia 2012)	GII.P7_GII.7 KJ196295 (Japan 2010)	Figure 5.5
GII.P16 ORF1	1 GII.P16_II.17 3 GII.P16_II.4	Maximum Likelihood	GII.P16_GII.2 KJ407074 (USA 2011) GII.P16_GII.17 KJ196286 (Japan 2002)	GII.P7_GII.7 KJ196295 (Japan 2010)	Figure **5.4**

Tree details	Number of sequences in tree*	Phylogeny method	Reference sequence	Out-group sequence	Figure No.
GII.P21 ORF1	64 GII.P21_II.3 1 GII.P21_II.13	Maximum Likelihood	GII.P21_GII.3 LN854569 (Netherlands 2014)	GII.P7_GII.7 KJ196295 (Japan 2010)	Figure 5.6
GII.P7 ORF1	13 GII.P7_II.6 3 GII.P7_II.7	Maximum Likelihood	GII.P7_GII.6 LN854568 (Netherlands 2014) GII.P7_GII.7 KJ196295 (Japan 2010)	GII.P12_GII.3 GU980585 (Korea 2006)	Figure 5.3
Longitudinal samples** full genome**	83	Maximum Likelihood	GII.Pc_GII.2 JX846925 (Kuala Lumpur 1978) GII.P21_GII.3 LN854569 (Netherlands 2014) GII.P4_GII.4 New Orleans_2009 JN595867 (USA 2010) GII.Pe_GII.4 Sydney_2012 JX459908 (Australia 2012) GII.P16_GII.17 KJ196286 (Japan 2002) GII.P17_GII.17 LC037415 (Kawasaki 2015) GII.P7_GII.6 LN854568 (Netherlands 2014)	GI.P3_GI.3 KJ196292 (Japan 2007)	Figure 5.26
Ribavirin treated Patient A and B VP1	6	Maximum Likelihood	None	GII.P.4_GII.4 Farmington Hills_2002	Figure 6.10

* excluding reference sequences

** Two samples excluded from phylogeny due to mixed genotypes

*** One sample excluded from phylogeny due to mixed genotypes

**** Three samples excluded from phylogeny; two due to mixed genotypes; one sample excluded (Px 101) as only one sample remaining once mixed sequence excluded

† Only 1 sample of this genotype, therefore phylogeny was not reconstructed

†† P2; capsid protruding domain 2, nt 5910–6336 (numbering corresponds to GII.Pe_GII.4 Sydney_2012 JX459908 reference sequence)

N/A not applicable

2.15 CLONING OF NS7 AND VP1 PCR AMPLICONS (CHAPTER 6)

VP1 and NS7 amplicons were cloned using a pGEM-T Easy Vector System (Promega) and α-Select chemically competent cells (Bioline), according to manufacturer's instructions. For each sample, 96 colonies were sub-cultured (into a 96-well plate, each well containing agar); DNA extraction and capillary sequencing of each sub-cultured colony was done by the GATC Biotech sequencing service.

2.16 DEEP SEQUENCING OF NS7 AND VP1 PCR AMPLICONS (CHAPTER 6)

NS7 and VP1 amplicons from each sample were pooled in equimolar quantities and mechanically sheared to 200 bp using a Covaris E210 focused-ultrasonicator (duty cycle 10%, Intensity 5 and 200 cycles per burst, 6 cycle repeats of 60 seconds). Sheared samples were prepared for deep sequencing using NEBNext DNA Library prep master mix set for Illumina and NEBNext Multiplex oligos for Illumina (New England Biolabs), according to manufacturer's instructions. Prior to sequencing the library quality was assessed using Bioanalyser HS chips (Agilent). Indexed samples were multiplexed prior to 250 bp paired end sequencing on an Illumina MiSeq platform.

2.17 CALCULATION OF MUTATION FREQUENCY IN NS7 AND VP1 AMPLICONS (CHAPTER 6)

Reference guided assembly was performed with Seqman NGen version 11 (DNA Star) using default parameters and a simple percentage SNP calculation method. SNP reports were generated by Seqman Pro version 11.1 (DNA Star). The consensus sequence of the pre-ribavirin sample for each patient, obtained by capillary sequencing, was used as the assembly reference. SNPs with frequency <1% and sites with a consensus sequence change compared to the reference were excluded from analysis.

Mutation frequencies of the total virus population were calculated as previously described [158], using the following equation;

$$No.\ mutations\ per\ 10{,}000\ nucleotides = \frac{No.\ SNPs}{Total\ No.\ bases\ sequenced} \times 10{,}000$$

2.18 CALCULATION OF dN/dS RATIOS (CHAPTER 6)

The degenerate nature of the amino acid code means that multiple codon sequences can code for the same amino acid. Consequently nucleotide substitutions (or mutations) in protein-coding genes can either be synonymous, in which case they do not result in an amino acid change (also known as silent substitutions), or non-synonymous, in which case they change the amino acid.

Under purifying selection, changes in a protein are disadvantageous therefore are selected against. Conversely under positive selection, changes in a protein are advantageous therefore are selected for. Synonymous substitutions (dS) do not have any impact on the protein, therefore are unlikely to be subject to purifying or positive selection; they are considered to be under neutral selection. Synonymous substitutions (dS) are an indication of background genetic drift.

The number of non-synonymous substitutions (dN) can be normalised against the number of synonymous substitutions to determine whether the protein sequence is under purifying, neutral of positive selection. This is known as the dN/dS ratio. The equation is laid out below, where dN is the number of non-synonymous substitutions and dS the number of synonymous substitutions. A dN/dS ratio of 1 indicates no selection is occurring (neutral), <1 suggests purifying selection and >1 suggests positive selection.

$$dN/dS\ ratio = \frac{dN}{dS}$$

2.19 MAPPING AMINO ACID RESIDUES IN THE NOROVIRUS PROTRUDING DOMAIN PROTEIN STRUCTURE (CHAPTER 6)

Amino acid sequences of interest were aligned with the P domain sequence of norovirus GII.4 VA387 (2OBT; P domain protomer and HBGA B trisaccharide complex [35]) using ClustalW in Megalign (DNA Star) v 12.2. The crystal structure was viewed in PyMol (v 1.7.4.4 Edu) and known epitopes identified as per **Table 2.4**).

Table 2.4 Known epitopes in the norovirus GII.4 protruding (P) domain

Epitope	Amino acids*	Evidence	Reference
P1 subdomain	222–274, 418–539	Crystal structure†	[35]
P2 subdomain	275–417	Crystal structure†	[35]
HBGA binding site	343–345, 374, 441–443 ††	Crystal structure†	[35]
Indirect role in HBGA binding (e.g. interaction with binding site)	336, 338, 347	Crystal structure†	[35]
Open pocket, speculated to interact with and increase affinity between trisaccharide and P protein	390–393, 395	Crystal structure†	[35]
Dimer interface	231, 232, 238, 243, 245, 274, 278–281, 308, 333, 337, 344, 346, 348, 384, 386, 436, 440–442, 455, 458, 459, 463	Crystal structure†	[35]
Binding pocket	291, 300, 335, 368	Sequence modelling and VLP ELISAs (mutations in VLP affected binding to HBGAs from saliva) Doubtful significance as not seen in crystal structure [35]	[159]
Immunogenic epitope A	294, 296–298, 368, 372	Changes correlate with shifts in GII.4 antigenicity	[123]
Immunogenic epitope D	393–395	Changes correlate with shifts in GII.4 antigenicity	[124]
Immunogenic Epitope E	407, 412–413	Changes correlate with shifts in GII.4 antigenicity	[124]

* amino acid position relative to GII.4 VA387 capsid sequence (Genbank accession number AAK84679);

†Crystalized P domain protomer and HBGA trisaccharide complex (2OBT, www.rcsb.org);

†† It is predicted there are additional interactions and binding sites with lower affinity not observed in complex crystals [35].

HBGA, human blood group antigen; VLP, virus like particle

2.20 STATISTICAL ANALYSIS

All statistical analysis was performed using IBM SPSS Statistics v23 using two-tailed tests at the 95% significance level.

2.20.1 Statistics for Chapter 3

The difference in age and PCR Ct values between gastroenteric viruses and between norovirus genotypes was tested using Kruskal Wallis ANOVA. Significant outcomes from Kruskal Wallace ANOVA were further tested by Pairwise comparison; P values were adjusted for multiple comparisons. The difference in age and PCR Ct values between norovirus genogroups and between healthcare/previously acquired infections (HAI/POA) was tested using a Mann Whitney U Test. The difference in proportion of each norovirus genotype between HAI/POAs and between clinical specialties was tested by Fishers Exact test. The relationship between new norovirus episodes per month and the number of admissions per month was tested by Pearson's correlation and Poisson Regression analysis.

The difference in proportion of norovirus chronically infected patients between clinical specialties and genotypes was tested by Fishers Exact test. The difference between duration of infection in chronically infected patients between clinical specialties and genotypes was tested using Kruskal Wallis ANOVA.

2.20.2 Statistics for Chapter 4

The difference in % on-target-reads (% OTR), read depth and % genome coverage between norovirus genotypes and in PCR Ct value between Pass/Sub-optimal/Failed samples was tested by Kruskal-Wallis ANOVA, with pairwise multiple comparison of significant results and P values adjusted for multiple comparisons.

The relationship between PCR Ct value and % OTR, read depth and % genome coverage was assessed by Spearman's correlation.

A simple linear regression model (independent variable, PCR Ct value; dependant variable, logit transformed %genome coverage) was fitted to generate prediction intervals for % genome coverage from the PCR Ct value. % genome coverage was transformed using the formula $tr\_genome = \frac{\%genome\ coverage \times (N-1)+0.5}{N}$ to ensure there are no proportions of 0 or 1 (SPSS cannot logit transform values of 0 or 1) and then transformed again (to enable linear regression) using the logit function ($logit\ transformed\ \%genome\ coverage = \log\left(\frac{tr\_genome}{1-tr\_genome}\right)$ where log is the natural logarithm with base e.

Twenty-one of 507 samples were considered outliers and therefore excluded from the linear regression model to predict % genome coverage. These samples were excluded following discussion with the ICH Statistical Support Service (support detailed in

section 4.5 on page 141); 3/21 were believed to be RNA extraction failures (detailed on page 126, **Table 4.2** and circled in **Figure 4.8**) and 18/21 were believed to be cDNA synthesis failures from Run 30 and 31 (detailed on page 125–126).

2.20.3 Statistics for Chapter 5

The difference in duration of infection between chronically infected patients who were and were not re-infected with a different strain of norovirus was assessed using a two-tailed Mann Whitney U Test with a 95% significance level.

2.20.4 Statistics for Chapter 6

The difference in mutation frequency between the capsid (VP1) and RdRp (NS7) regions in all patients was assessed using a Wilcoxon signed rank paired test, and the difference in mutation frequency between pre- and post-ribavirin samples in Patient A and Patient B was tested using a Mann Whitney U Test. All tests were two-tailed with a 95% significance level, using transition frequencies only.

CHAPTER 3

PREVALENCE OF NOROVIRUS INFECTION AND GENOTYPES IN A PAEDIATRIC UK HOSPITAL WITH IMMUNOCOMPROMISED PATIENTS

3.1 INTRODUCTION

Norovirus is a leading cause of viral gastroenteritis world-wide [107]; however acute gastroenteritis can also be caused by adenovirus, astrovirus, sapovirus and rotavirus.

Adenoviruses are the only of the gastroenteric virus with a double stranded DNA genome, in the range of 33–45 kbp. Human adenoviruses (species names *Human mastadenovirus A–F*) belong to the genus *Mastadenovirus* in the *Adenoviridae* family. Adenoviruses are a broad group of viruses, both in terms of genome sequences and clinical syndromes. There are 47 serotypes described each with different disease associations; the most common being respiratory disease, conjunctivitis and gastroenteritis. Acute gastroenteritis is caused by only two serotypes, 40 and 41. In immunocompetent individuals adenovirus infections are mostly mild and self-limiting, however following primary infection they can persist in a latent state [160]. The predominant cause of adenovirus-associated disease in immunocompromised patients is from reactivation of latent infections, in which detection in stool can be used for early initiation of antiviral treatment to prevent progression to life-threatening invasive infection [161].

Astrovirus and sapoviruses are both small single stranded RNA viruses, with genome size of 6.2–7.8 and 7.1–7.7 kb, respectively. Sapoviruses (species *Sapporo virus*) are the most similar to norovirus; as they belong to the *Sapovirus* family in the genus *Caliciviridae* they have similar cupped surface morphology typical of the *Caliciviridae*. *Human astrovirus* belong to the *Mamastrovirus* genus in the family *Astroviridae.* Outbreaks caused by sapoviruses occur throughout the year, less frequently than for noroviruses [162]. Astrovirus infections typically only occur in children, with very few reports in adults.

Rotaviruses, in the *Rotavirus* genus, belong in the family *Reoviridae* (sub-family *Sedoreovirinae*) and have a segmented double stranded RNA genome. The vast majority of human rotavirus infections are caused by *Rotavirus A*, with occasional episodes of gastroenteritis caused by *Rotavirus B* and *Rotavirus C*; consequently it is generally only *Rotavirus A* which is the target for diagnostic PCR assays. Rotavirus infections are typically the most clinically severe of the gastrointestinal viruses, accounting for 70% of hospitalisations for gastroenteritis during seasonal peaks [163]. Consequently in 2013 a two-dose live oral vaccine was introduced to the UK childhood immunisation schedule with a dose at 2 months and 3 months of age. In the UK rotavirus infections have reduced by 67% since the introduction of the vaccine [164], however it has been shown that 17% of children shed vaccine virus in their stool for up

to 8 months following vaccination [165]. The vaccine is contraindicated in immunosuppressed patients; in the event that a child is vaccinated before diagnosis of their immunodeficiency, vaccine virus can cause a symptomatic infection [166, 167].

In 2010 the prevalence of healthcare-associated viral gastroenteritis in a paediatric UK hospital was assessed by Cunliffe *et al*. [168], who detected a viral cause in 53% of patients. Of 576 episodes of laboratory confirmed gastroenteritis, the majority (31%) were caused by rotavirus, an equal number (16% and 15%) identified norovirus and adenovirus, respectively, and a minority (5% and 2%) of infections were caused by astrovirus and sapovirus, respectively. This study was conducted prior to the implementation of the rotavirus vaccine; it is not known whether the prevalence of non-rotavirus gastrointestinal viruses has changed in the relative absence of rotavirus infections.

Although GII.4 strains predominate in norovirus outbreaks, they do so in the context of co-circulating genotypes [129]. Earlier reports have suggested that some norovirus genotypes are more common in children than adults [169, 170]; anecdotally this is particularly the case for GII.3 norovirus however there is limited evidence of this in the literature. The distribution of genotypes in a UK nosocomial paediatric population has not previously been described. In addition, it is not known whether chronic norovirus infections in immunocompromised patients are associated with specific genotypes.

The site of norovirus *in vivo* replication remains elusive. Recent data from Jones *et al.*[24] reported *in vitro* culture of human norovirus in B cell lines, suggesting that B cells could be a major target for norovirus replication *in vivo.* At GOSH we are involved in the management of between 40–50 children with severe combined immunodeficiency (SCID). These patients are born with genetic defects affecting T and B lymphocyte development and function, although the degree to which each lineage is affected varies by mutation. Some of these patients do not have B cells; one would expect that if B cells are the primary site of norovirus replication, these patients could be less susceptible to norovirus infection compared to patients with B cells.

This chapter describes the prevalence of norovirus in a paediatric UK hospital, in the context of other gastrointestinal infections occurring in this population. It explores the relationship of different norovirus genotypes and other gastrointestinal viruses to seasonality and viral load. Understanding the molecular epidemiology of viral gastroenteritis in children will contribute to improving infection control practices and vaccine development. In addition we evaluated children with and without B lymphocyte-

deficient SCID to determine whether the former were protected against norovirus infection.

3.2 MATERIALS AND METHODS

3.2.1 Sampled population

Between 1st July 2014 and 30th June 2015 a total of 4,786 stool samples from 1,393 patients (8% outpatients) at GOSH were tested during routine diagnostic analyses for the presence of gastroenteric viruses (**Table 3.1**). Between 1 and 46 samples were tested per patient (median 1) (**Figure 3.1**). The norovirus PCR assay was developed as part of this study (all other real-time PCR assays were developed by the GOSH Virology department). Detailed methods are described in Chapter 2.

Table 3.1 Demographics of 4,786 stool samples tested in this study between 1/07/2014 and 30/06/2015 at GOSH, UK

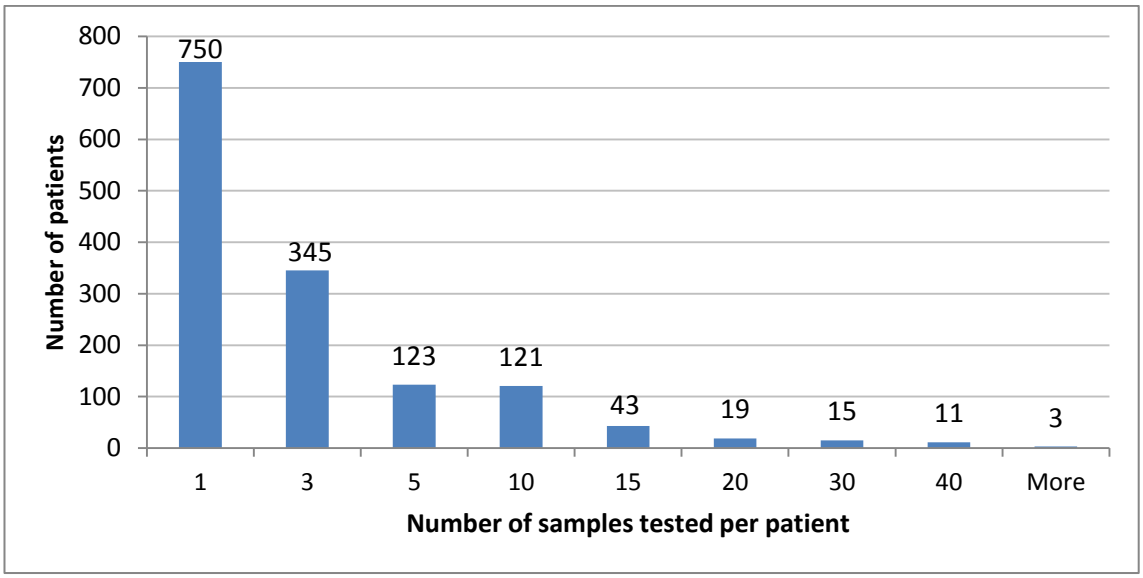
Clinical specialty	Number of patients tested (% male)	Median age (range), years
Surgical	202 (55%)	1.5 (0–17)
Medical*	803 (54%)	1.9 (0–18)
Immunocompromised†	270 (56%)	2.7 (0–17)
International and Private‡	118 (59%)	3.3 (0–15)
TOTAL	**1393 (55%)**	**2.1 (0–18)**

* Medical; respiratory medicine, cardiac medicine, renal medicine, intensive care, neurology, dermatology, rheumatology, ear nose and throat and ophthalmology

† Profoundly immunocompromised; bone marrow transplant, oncology, haematology and immunology

‡International and Private; range of clinical specialties

Figure 3.1 Number of stool samples tested per patient, July 2014–June 2015, for detection of gastrointestinal viruses at GOSH (UK)



Diagnostic real-time PCR results and accompanying clinical data for this study were exported retrospectively from the laboratory information system. The vaccine status of individual patients is not known. Detection of more than one virus during the study period was treated as an independent episode in the analysis.

3.2.2 Norovirus genotyping

The first norovirus positive stool from each patient at GOSH between 1st July 2014 and 30th June 2015 was prospectively genotyped by PCR amplification and capillary sequencing of the capsid shell domain, described in detail in Chapter 2. Eleven norovirus positive samples had insufficient residual volume for genotyping and were excluded from analysis. cDNA for genotyping was prepared from 133 samples.

3.2.3 National norovirus genotyping data

To compare norovirus seasonality and genotypes circulating at GOSH to those circulating nationally, the number of norovirus outbreaks reported nationally to Public Health England (PHE) and the proportion of each genotype in the laboratory-confirmed outbreaks was provided by the Virus Reference Department (VRD), PHE, from their national surveillance data (collected via the Hospital Norovirus Outbreak Reporting System and published monthly on https://www.gov.uk/government/statistics/norovirus-national-update).

3.2.4 Rotavirus vaccine detection

Rotavirus positive specimens were genotyped, and GIP8 positives confirmed as vaccine or wild-type by sequencing the genes encoding VP4 and VP7, by the PHE VRD.

3.2.5 Categorisation of patients

The clinical specialty of each patient was assigned based on the clinical specialty of the ward to which they were admitted at the time of specimen collection. Immunodeficiency patients consisted of specialties associated with profound immunodeficiency; bone marrow transplant, oncology, haematology and immunology specialties. Medical patients consisted of respiratory medicine, cardiac medicine, renal medicine, intensive care, neurology, dermatology, rheumatology, ear nose and throat and ophthalmology. It is possible that some patients in the medical category will have had some degree of suppressed immunity.

3.2.6 Statistical analysis

Statistical analysis was performed using IBM SPSS Statistics v23. Details are described in Chapter 2.

3.2.7 Norovirus in B cell deficient patients

SCID patients are a very rare cohort, therefore to maximise numbers we retrospectively reviewed the records from 2008 onwards for children with T−B−NK+ SCID in whom B and T cells were absent. These included five with Rag-1 mutations resulting in deficiency of RAG 1 (recombination-activating gene type 1) protein (Px 1–4 and Px 6) and two with DCLRE1C mutations resulting in Artemis deficiency (Px 5 and Px 8). We compared these with seven children designated T−B+NK− SCID in whom mutations in IL2RG resulted in complete absence of T cells and natural killer (NK) cells with present, albeit dysfunctional, B lymphocytes (Px 11–13 and Px 15–18), one T−B+NK+ child in whom mutations in IL7RA results in absence of T cells only (Px 10) and one each RAG 1 (Px 14) and Artemis deficiency (Px 9) patients who had undergone hematopoietic stem cell transplant (HSCT) thus introducing B cells.

The difference in norovirus viral titre between B− and B+ patients was assessed by comparing the Ct values of all norovirus positive samples from each patient using a two-sided two sample T-test, with a 95% significance level.

B cell counts in whole blood are measured by the diagnostic Immunology department. Normal B cell counts in children 2 months to 15 years are 0.6–2.7 $x10^9$ cells/L [171].

3.3 RESULTS

3.3.1 Validation of performance characteristics of norovirus real-time PCR

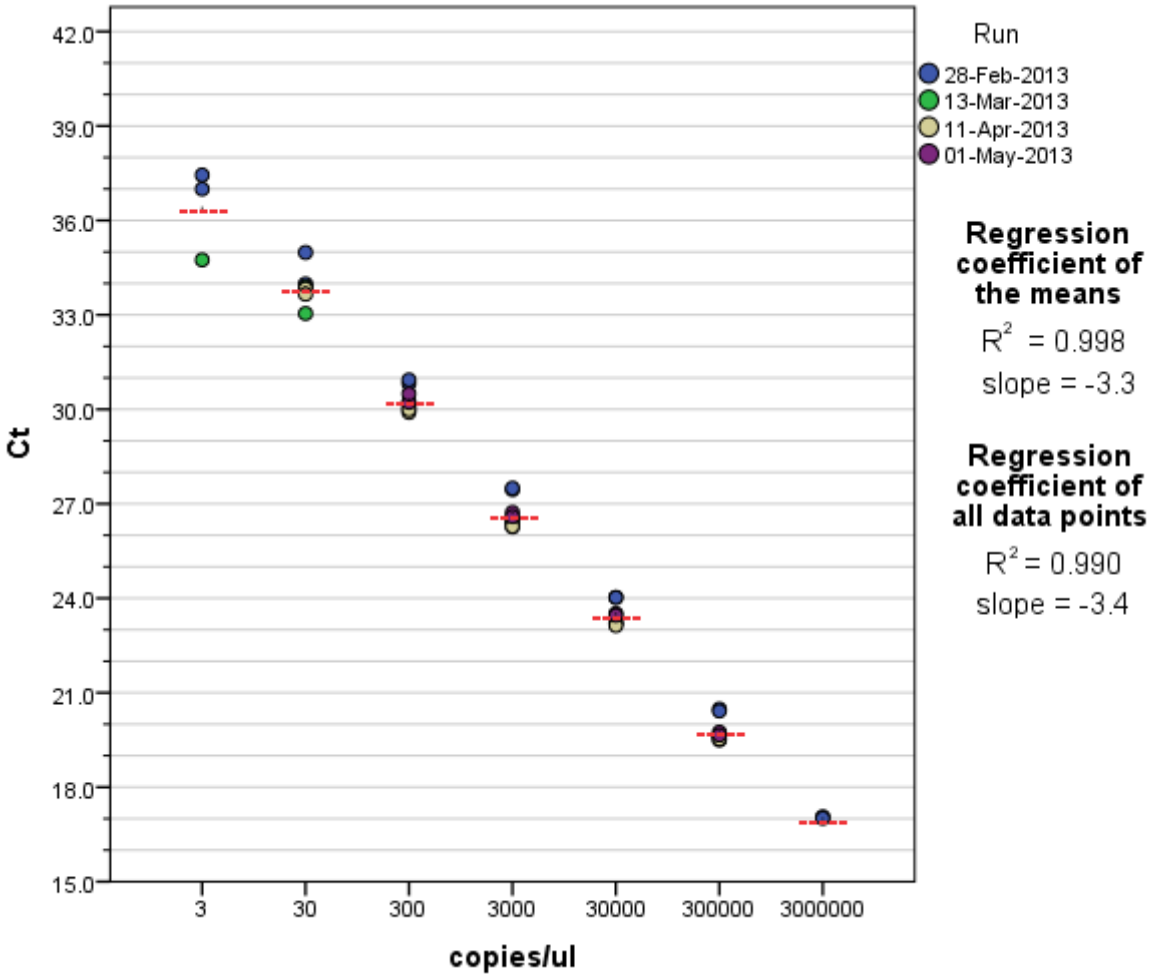
A nucleotide BLAST search of the norovirus GI and GII primers and probes only returned matches against norovirus GI or GII, respectively. Adenovirus, rotavirus, sapovirus and astrovirus were not amplified by the norovirus RT-qPCR assay when tested on stool samples known to be positive for one of the aforementioned viruses. These results confirm the primer and probes are specific to norovirus.

The lowest concentration of norovirus plasmid detected, and therefore the limit of detection, was 5 and 1 copies/µl for GI and GII, respectively.

All of the GI (4/4) and GII (5/5) positive samples from QCMD were amplified and correctly identified as GI or GII. The GIV positive sample was amplified by the GI primers and probe.

A linear relationship is observed between norovirus viral load in copies/µl and Ct value, with a difference in 3.3 Ct values for every 10-fold change in viral load. The variability in Ct value between different runs was minimal (± 1 Ct, or 0.25 log10) (**Figure 3.2**).

Figure 3.2 Real-time PCR Ct values of known quantity plasmid containing norovirus GII target sequence, tested on multiple days. Red lines indicate mean Ct value



Copies/µl	3	30	300	3000	30000	300000	3000000
N	3	6	8	8	7	8	2
Mean Ct	36.4	33.9	30.3	26.7	23.5	19.8	17.0
Standard deviation	1.4	0.6	0.4	0.5	0.4	0.4	0.03
Standard error	0.8	0.3	0.1	0.2	0.1	0.1	0.02

3.3.2 Prevalence of gastroenteric viruses

Twenty-four percent (329/1393) of all patients tested in the twelve month period were positive for a gastroenteric virus, among which norovirus and adenovirus predominated with 144 and 146 episodes each over the 12 month period, each constituting 44% of all viral gastrointestinal infections (**Table 3.2**).

Table 3.2 Number and proportion of real-time RT-PCR positive episodes at GOSH, a children's tertiary referral hospital, UK, over 12 months (2014–2015) from 4,786 stool specimens

	Total number	% of patients	% of positive samples	Mixed infections (% of positives)	Chronic infections† (% of positive)
Number of patients tested	1,393				
Enteric virus detected	329	24%	100%	44 (13%)	
Norovirus	144	10%	44%	24 (17%)	38/144 (26%)
Adenovirus	146	10%	44%	28 (19%)	28/146 (19%)
Rotavirus	33 (8/29 vaccine*)	2%	10%	5 (15%)	6/33 (18%)
Sapovirus	80	6%	24%	17 (21%)	12/80 (15%)
Astrovirus	18	1%	5%	4 (22%)	1/18 (6%)

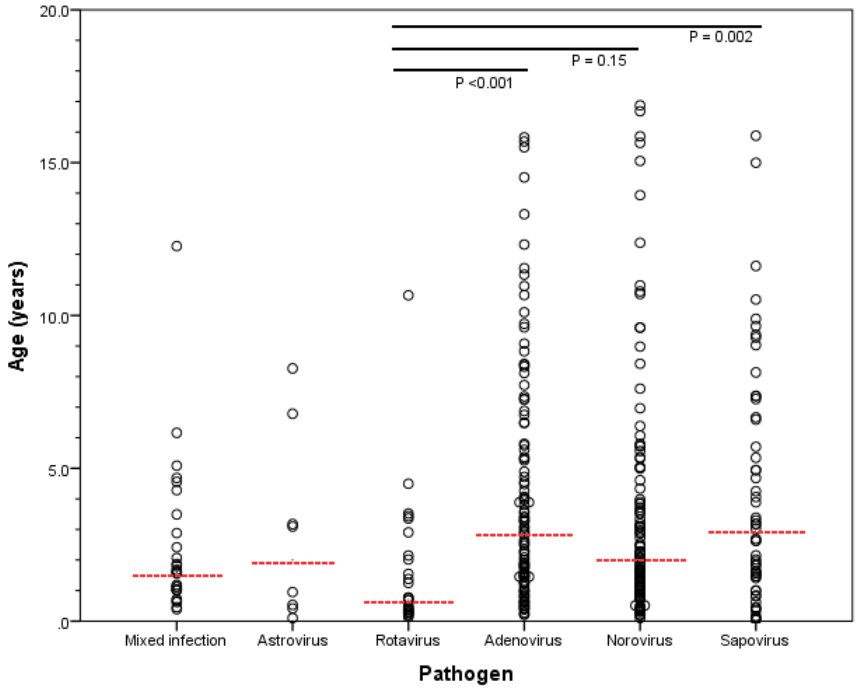
*29/33 rotavirus positive samples were typed for vaccine/wild-type; 4/33 samples had no residual specimen for typing

† Chronic infections based on PCR positive >1 month

Forty-four of the 329 infections (13%) were mixed infections with more than one virus detected (**Table 3.2**). The predominant mixed infections (23 /44, 52%) were norovirus and adenovirus, followed by norovirus and sapovirus (7/44, 16%) and adenovirus and sapovirus (6/44, 14%). Rotavirus was least frequently detected as part of a mixed infection (5/44, 11%). An equal number of mixed infections were from medical and immunocompromised patients; 16/44 (36%) each. Given that only 19% of patients tested were in the 'immunocompromised' category, this suggests mixed infections are likely to be more frequently associated with immune dysfunction.

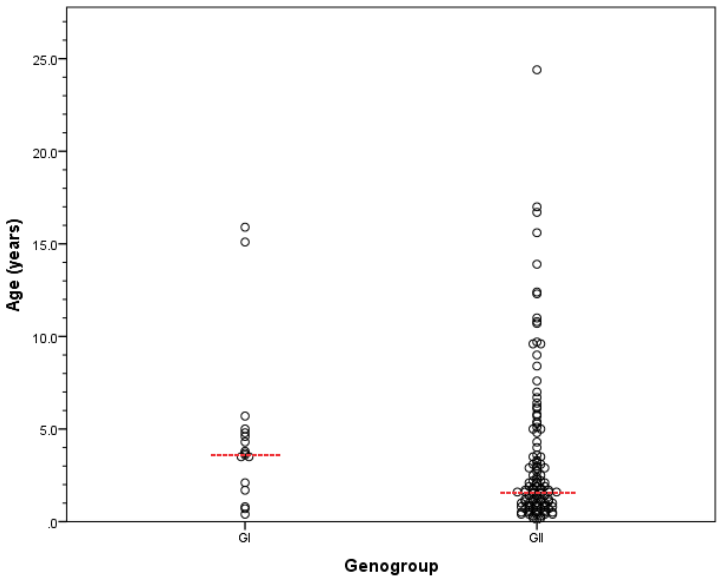
The median age of patients with a rotavirus infection, 0.7 years, was significantly younger than other infections with a median age of 2–3 years (P ≤0.015, **Figure 3.3**).

Figure 3.3 Age range of children infected with gastroenteric viruses. Horizontal lines indicate median age (n = 329)



Children with GI norovirus infections were on average older than GII infections (median age 3.7 and 1.7 years, respectively) however this difference did not reach significance (P = 0.061, **Figure 3.4**).

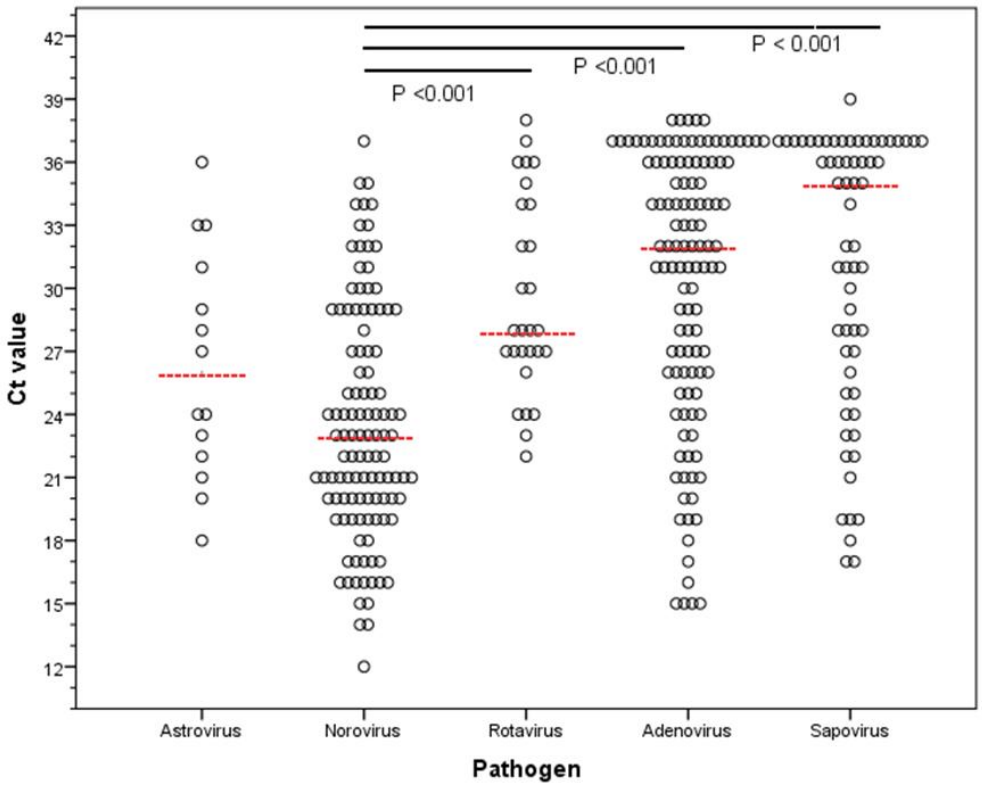
Figure 3.4 Age range of norovirus GI and GII infections in children (n = 133). Horizontal lines indicate median age



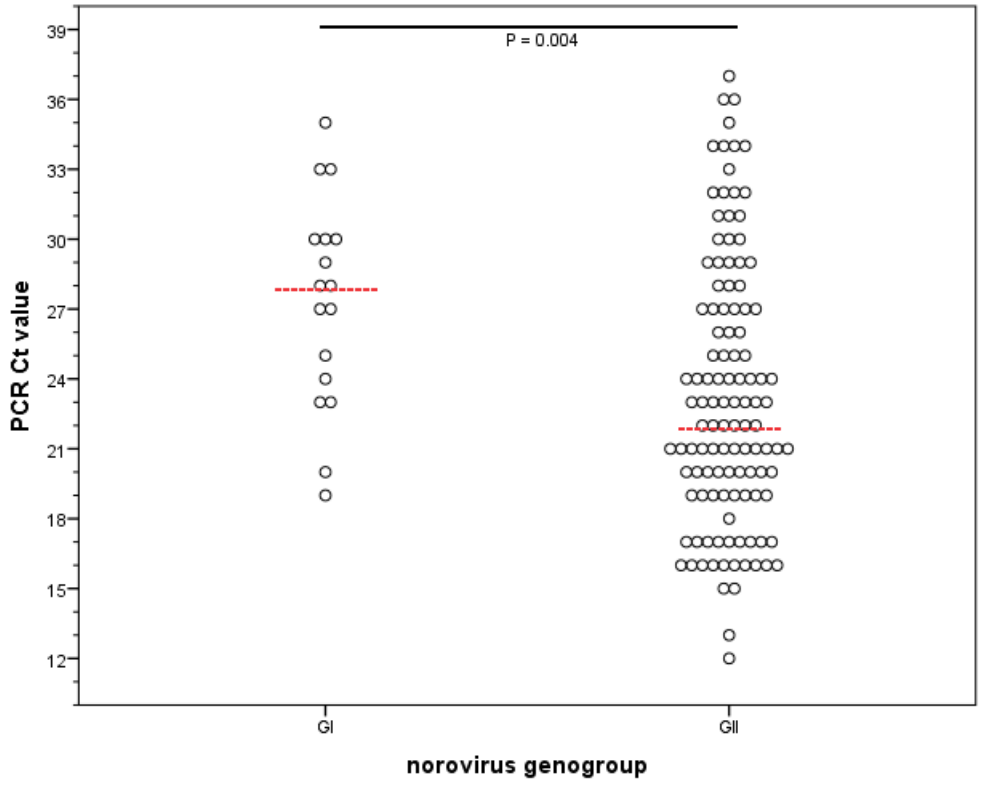
Norovirus infections had a significantly higher virus titre, median Ct 23, compared to other infections (P ≤ 0.03 **Figure 3.5a**). Sapovirus infections had the lowest viral titre; median Ct 35.

Figure 3.5 Real-time PCR Ct value range in children infected with **(a)** all gastrointestinal viruses (n = 329) and **(b)** norovirus GI or GII (n = 133). Horizontal lines indicate median PCR Ct value

(a)



(b)



3.3.3 Rotavirus vaccine-derived infections

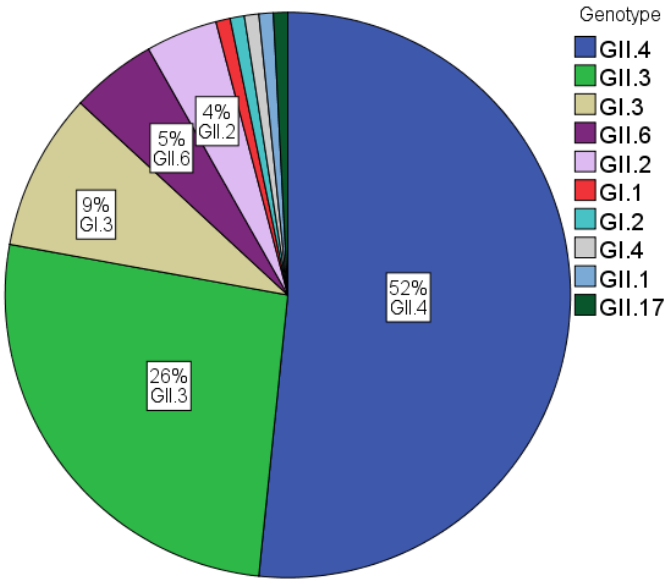
Four rotavirus positive patients had insufficient residual specimen for genotyping, thus 29 of 33 samples were genotyped. 28% (8/29) of rotavirus infections were identified as vaccine strain.

3.3.4 Prevalence of norovirus genotypes

Eighty-seven percent (117/133) of norovirus infections were genogroup II (GII), which had a significantly higher virus burden (median Ct 22) compared to genogroup I (GI) infections (median Ct 28) (P = 0.004, **Figure 3.5b**).

The majority of genotyped norovirus infections were GII.4 and GII.3; 52% and 26%, respectively (63/122 and 32/122), with the remaining 22% (27/122) identified as GI.1, GI.2, GI.3, GI.4, GII.1, GII.2, GII.6 or GII.17 (**Figure 3.6**). Eleven samples (11/133, 8%) could not be amplified by the genotyping PCR; these had a significantly lower viral burden compared to other samples (median Ct 35 and 22 for failed versus successful typing, respectively, P ≤ 0.001).

Figure 3.6 Proportion of norovirus genotypes detected in children over a 12 month period 2014–2015 (n = 122)



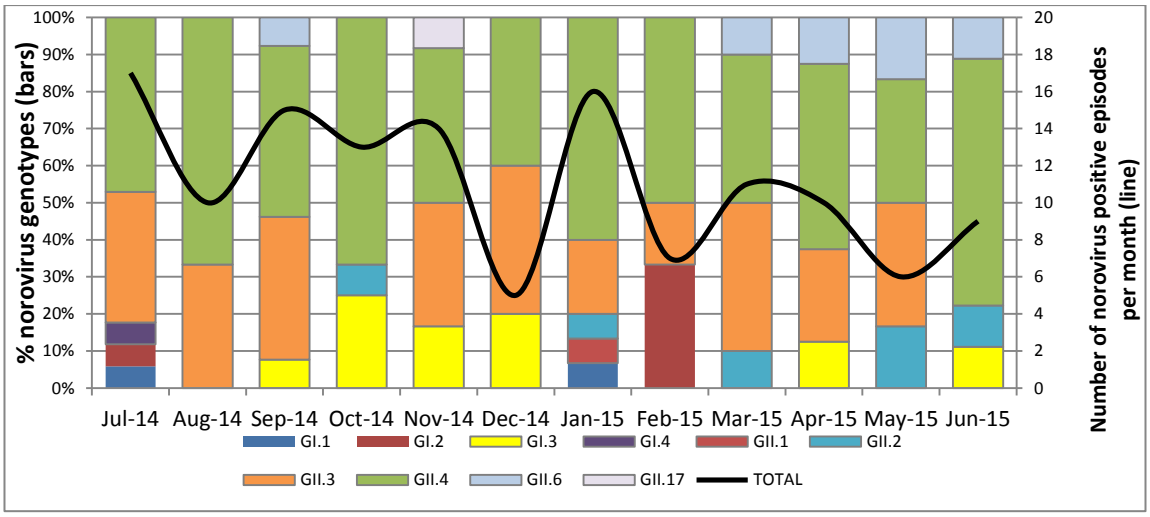
3.3.5 Norovirus seasonality

The proportion of norovirus genotypes each month in our paediatric population is not the same as those seen in nationally reported outbreaks, primarily attributable to the increased proportion of GII.3 in our population (**Figure 3.7**). A peak in prevalence of GI.3 in nationally reported outbreaks from August to November 2014 is followed by a

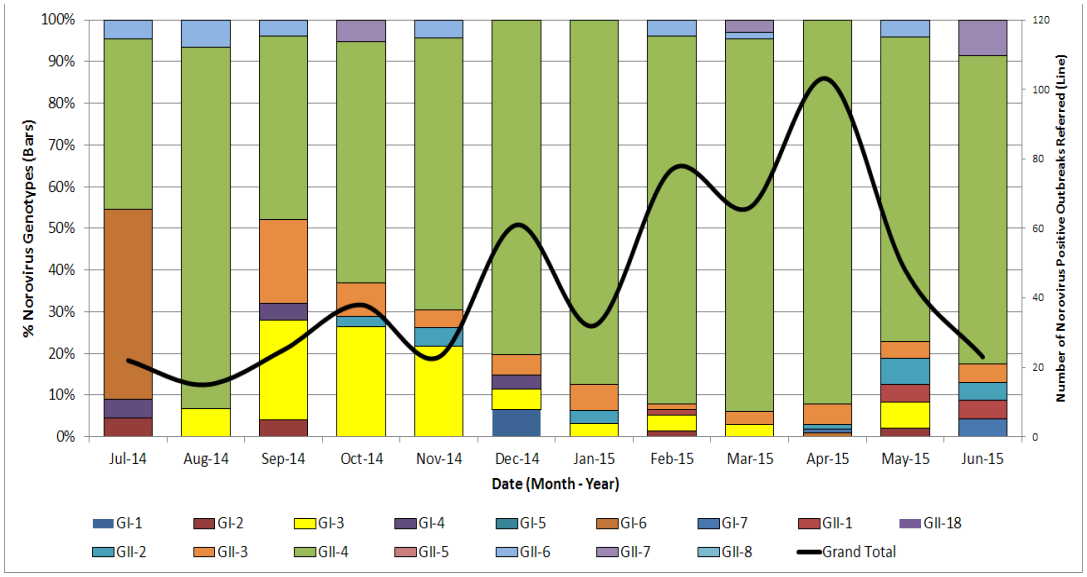
similar peak in our paediatric population from September to December 2014. Conversely, a peak in GII.6 episodes in our population from March to June 2015 is not seen in nationally reported outbreaks (**Figure 3.7**).

Figure 3.7 Proportion of norovirus genotypes by month in **(a)** children in a paediatric tertiary referral hospital (data from this study) and **(b)** nationally reported outbreaks, reproduced with permission from PHE norovirus surveillance report. Black lines indicate total number of outbreaks or cases per month

(a)



(b)

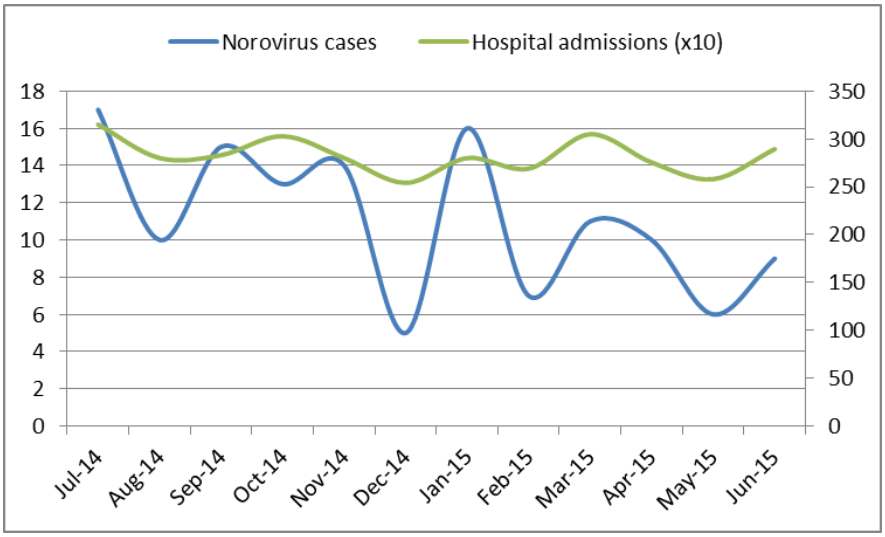


The overall number of norovirus cases per month in our population does not follow the typical winter peak seen in national outbreaks (**Figure 3.7**). Instead it was noted that in our population the number of cases of norovirus per month follows a similar trend to the number of hospital admissions, including outpatient visits and transfer between wards

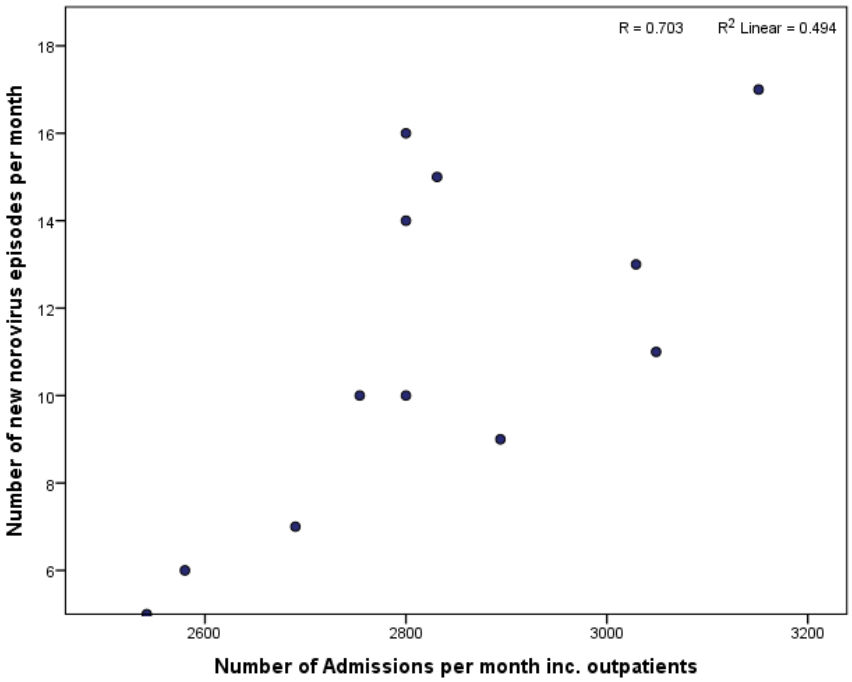
(**Figure 3.8**). There is a significant positive correlation between the number of admissions and number of norovirus cases per month (R = 0.703, P = 0.011). This suggests that the number of hospital admissions accounts for 50% of the variability in number of norovirus cases (R^2 = 0.494). Based on the Poisson regression coefficient ($y = -1.447 + 0.001x$) it is estimated that one case of norovirus occurs for every 100 admissions (95% CI 0.000–0.002, P = 0.002).

Figure 3.8 (a) Number of norovirus cases per month in paediatric tertiary referral hospital and number of hospital admissions during the same period. Admissions include outpatient appointments and transfers between wards; **(b)** Regression analysis of norovirus cases per month and number of hospital admissions (P = 0.006)

(a)



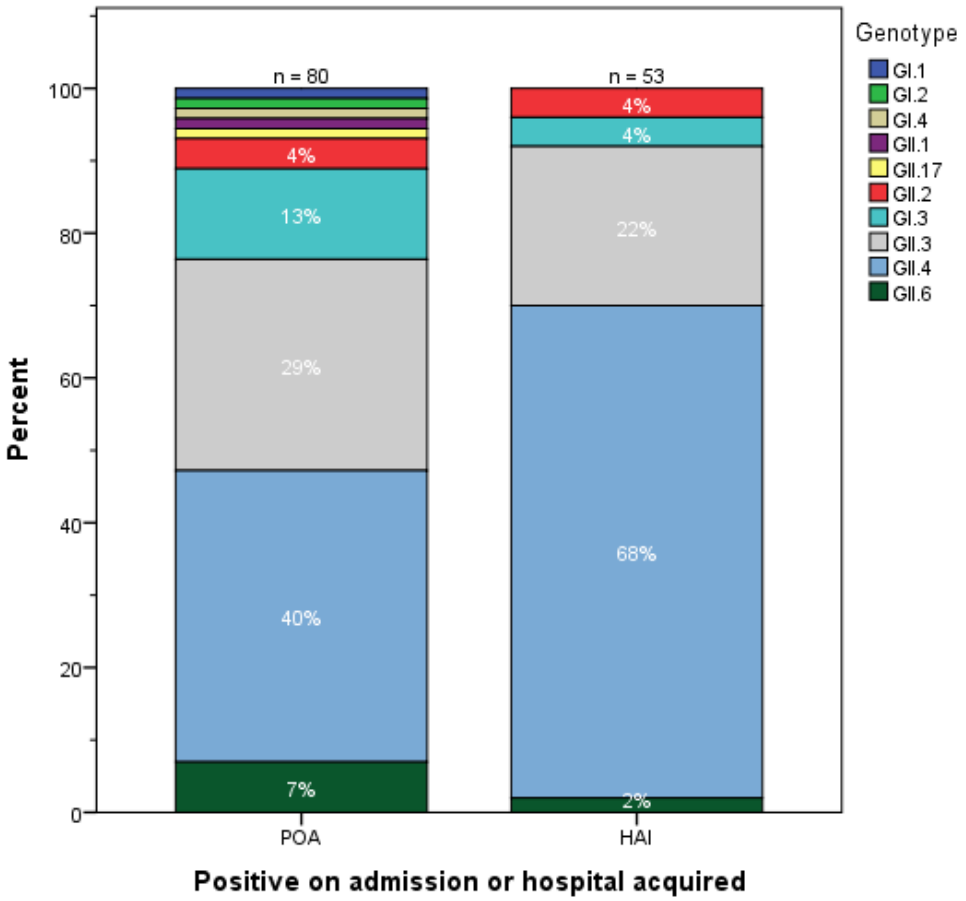
(b)



3.3.6 Hospital and community acquired norovirus infections

The proportion of GII.4 infections was 40% in infections acquired before admission (POA) and 68% in hospital acquired infections (HAI) (**Figure 3.9**); these were not significantly different (P = 0.062).

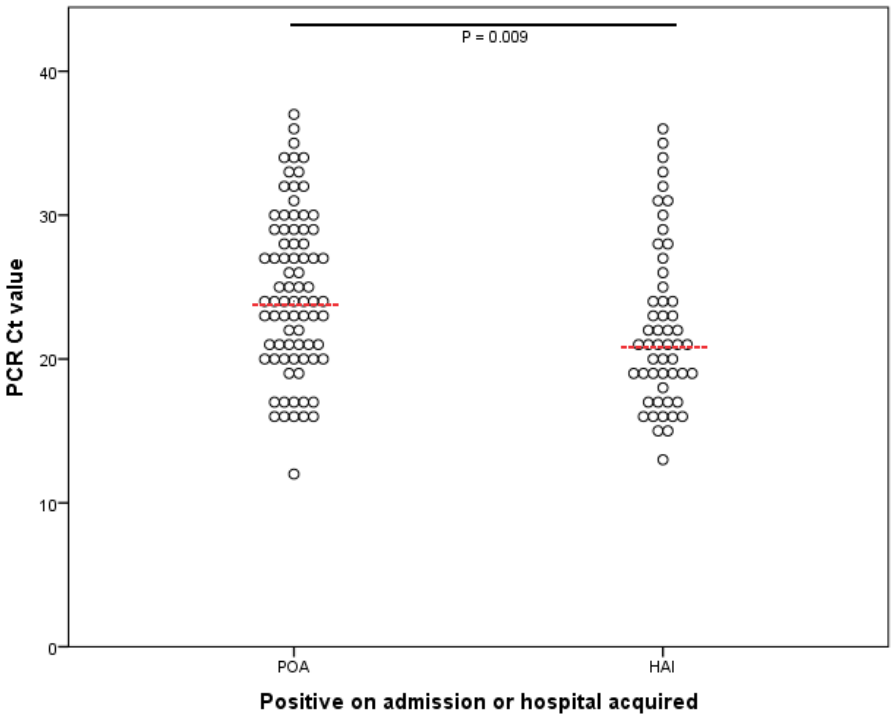
Figure 3.9 Proportion of norovirus genotypes in positive-on-admission (POA) and hospital acquired (HAI) norovirus infections



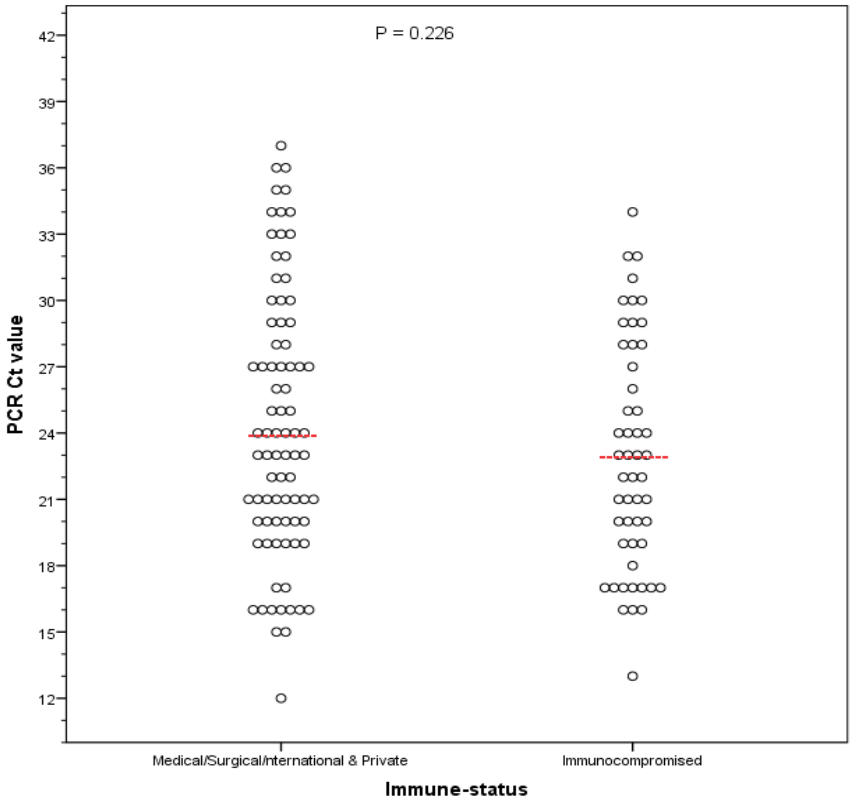
The average viral titre of norovirus hospital acquired infections (HAI) is significantly higher than for infections acquired before admission (POA), with a median Ct value of 21 and 24, respectively (P = 0.009, **Figure 3.10a**).

Figure 3.10 Real-time PCR Ct value range of **(a)** positive on admission (POA) and hospital acquired (HAI) norovirus infections (n = 133) and **(b)** norovirus infections in immunocompromised and non-immunocompromised patients. Horizontal lines indicate median Ct value

(a)



(b)



3.3.7 Norovirus in clinical specialties

The prevalence of norovirus infection is higher in immunocompromised compared to surgical or medical patients; 19% (51/270) of immunocompromised patients tested were found to be norovirus positive compared to 5% (10/202) and 7% (57/803) of surgical and medical patients.

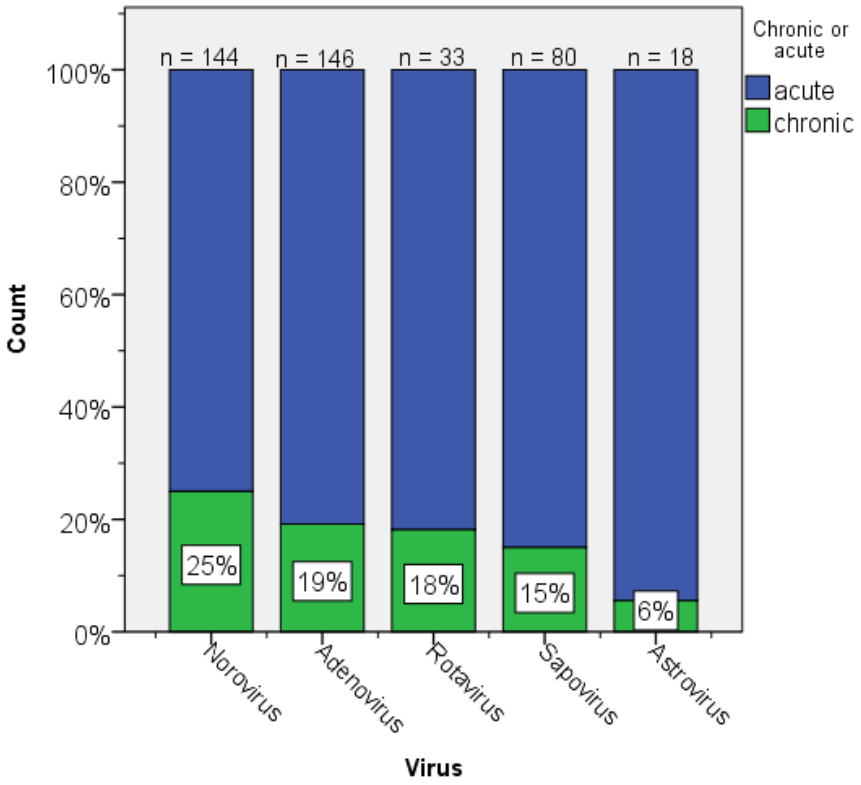
There was no significant difference in the norovirus PCR Ct values between immunocompromised and non-immunocompromised patients (median Ct 23 and 24, respectively; P=0.226, **Figure 3.10b**).

3.3.8 Chronic infections

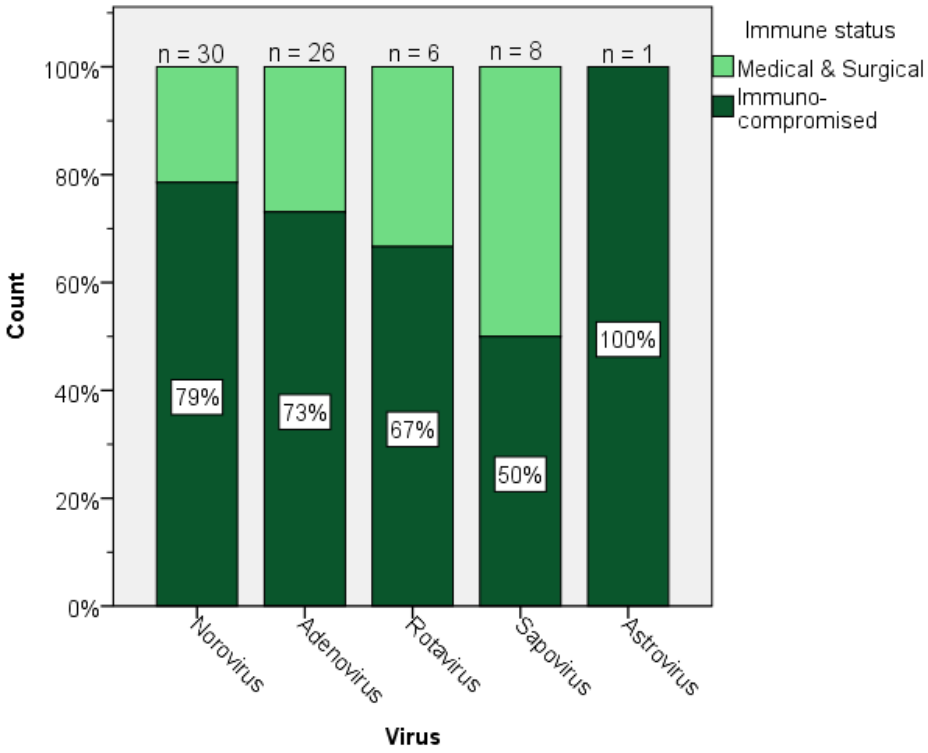
Norovirus had the highest rate of chronic infections (38/144, 25%); adenovirus, rotavirus and sapovirus had similar rates whilst astrovirus had the fewest (1/18, 6%) (**Table 3.2**, **Figure 3.11a**). With the exception of sapovirus, in which chronic infections occurred equally in immunocompromised and medical clinical specialties, the majority (67–100%) of chronic infections were in patients from immunocompromised clinical specialties (**Figure 3.11b**).

Figure 3.11 (a) Proportion of acute and chronic (PCR positive >1 month) infections and **(b)** Proportion of chronic infections belonging to immunocompromised or medical & surgical clinical specialties (excluding International & Private patients for whom immune status is not known*)

(a)



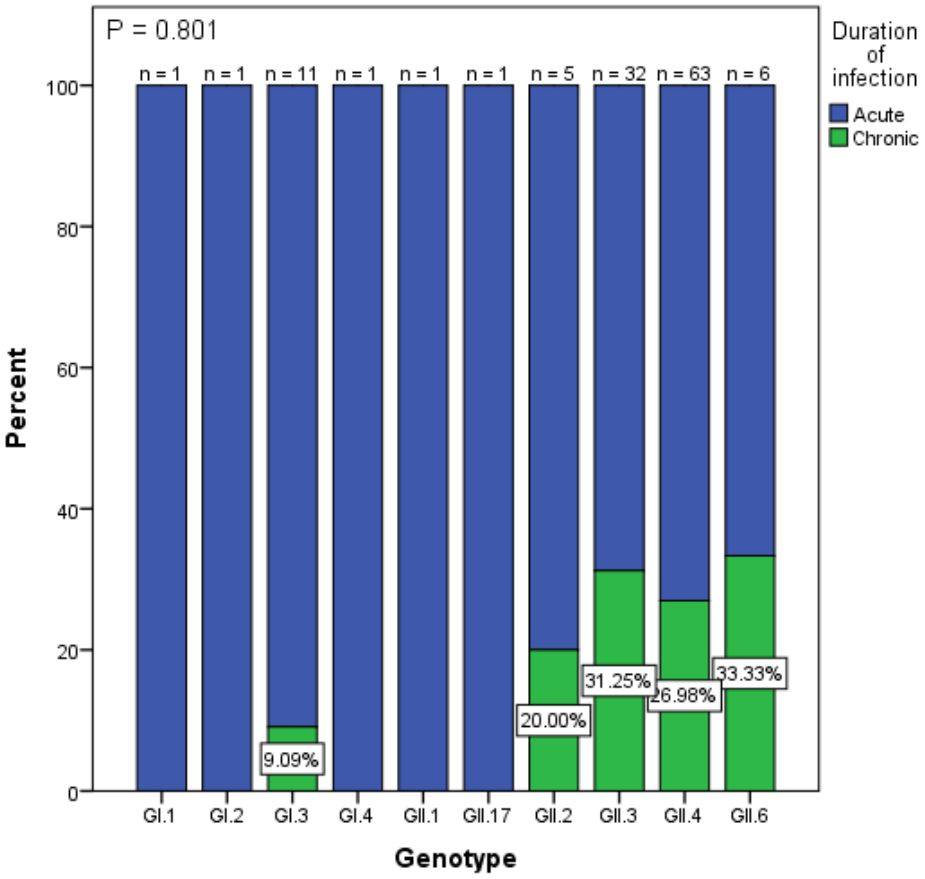
(b)



* Excluded chronically infected International & Private patients: norovirus 8/38; adenovirus 2/28; rotavirus 0/6; sapovirus 4/12; astrovirus 0/1

There was no difference in proportion of chronic patients between the different norovirus genotypes (**Figure 3.12**, P = 0.801). The median duration of infection in chronically infected patients was 5 months (range 1–21 months).

Figure 3.12 Proportion of norovirus genotypes with chronic infections



3.3.9 Norovirus in B cell deficient patients

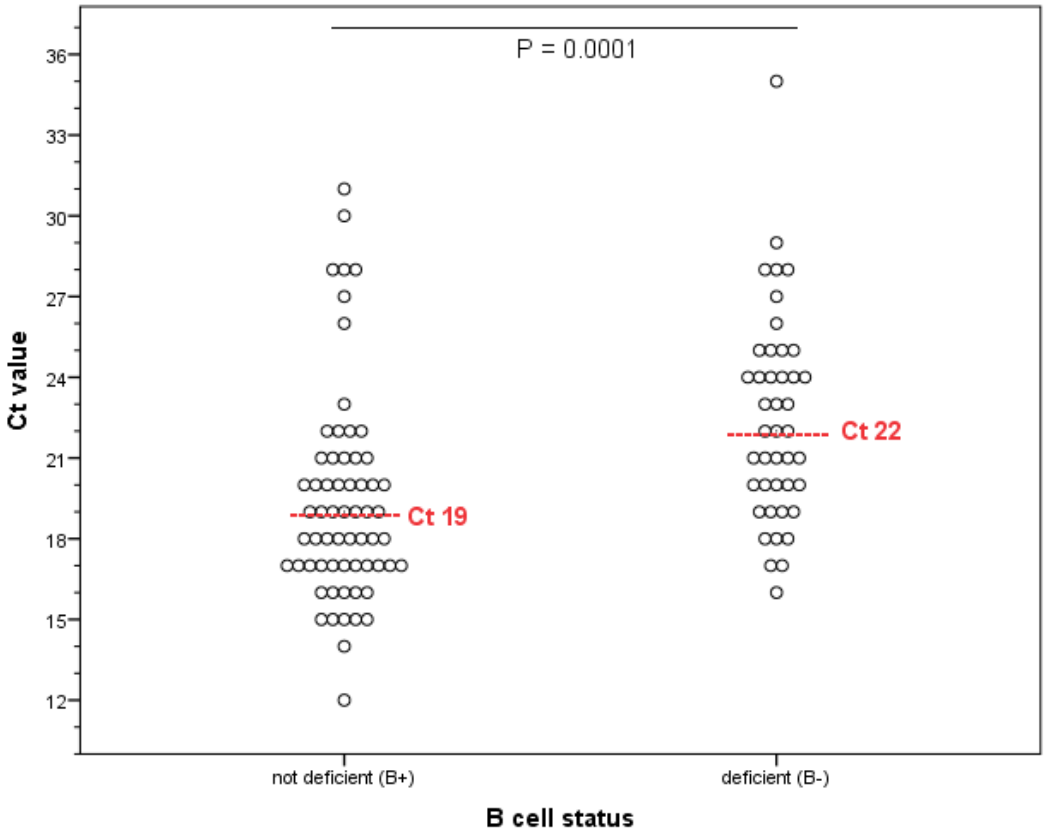
Of eight B− patients, five developed norovirus GII infection (63%), compared to 6 of 10 B+ patients (60%) (**Table 3.3**). The duration of virus shedding could not be determined for the majority of patients, however for the nine patients who had been diagnosed with norovirus infections since 2010, we were able to use Ct values to analyse viral titre. The average viral titre for B+ patients (mean Ct 19, 95% CI 18, 20) was a log higher than for B− patients (mean Ct 22, 95% CI 21, 23); this difference is significant (P 0.0001, **Figure 3.13**).

Table 3.3 Summary of norovirus results in B cell deficient and non-deficient SCID patients at Great Ormond Street Hospital, UK

Patient ID	Diagnosis	B cell status	B cell count (x10^9 cells/litre)	Norovirus infection	Median Ct† value	Ct value range	Number of positive samples detected	Date range of positive samples**
Px 1	RAG 1	−	0.04	yes	27	19-35	9	2.5 months
Px 2	RAG 1	−	0.04	yes	18	single result	1	single result
Px 3	RAG 1	−	0.01	yes	28	single result	1	single result
Px 4	RAG 1	−	0.00	yes	21	18-25	27	5 months
Px 5	Artemis deficiency	−	0.04	yes	17	16-22	5	14 months
Px 6	RAG 1	−	0.06	no				
Px 7	SCID RAG/Omenn	−	0.00	no				
Px 8	Artemis deficiency	−	0.01	no				
Px 9	Artemis deficiency	+*	0.29	yes	NA	NA		
Px 10	IL7RA	+	0.48	yes	NA	NA		
Px 11	X-SCID	+	2.36	yes	19	15-28	27	12 months
Px 12	X-SCID	+	2.84	Yes	22	21-28	7	10 months
Px 13	X-SCID	+	Not available	yes	17	12 to 22	25	12 months
Px 14	RAG 1	+*	0.69	yes	29	20-31	4	1 month
Px 15	X-SCID	+	1.52	no				
Px 16	X-SCID	+	2.13	no				
Px 17	X-SCID	+	1.03	no				
Px 18	X-SCID	+	0.93	no				

B–, B cell deficient; B+, B cell non-deficient; *, post-HSCT; NA, norovirus diagnosis was by electron microscopy therefore Ct values are not available; †, PCR cycle threshold. There is an inverse relationship between Ct and viral titre, with smaller values indicating a higher virus titre. A difference in Ct value of 3 is equivalent to a ten-fold difference in virus titre; ** date range of positive samples does not necessarily reflect duration of infection, as the dates of onset and end of infection were not known in all cases.

Figure 3.13 Real-time RT-PCR Ct value of all norovirus positive samples from B cell deficient (B−) and non-deficient (B+) patients. Lines indicate mean Ct value in each group



3.4 DISCUSSION

3.4.1 Norovirus is the most prevalent cause of gastroenteritis

This chapter presents the prevalence of viral gastrointestinal infections and the prevalence of norovirus genotypes in a large cohort of 1,393 paediatric patients in a tertiary referral hospital over a 12 month period, which is dominated by norovirus and adenovirus infections. This is similar to previous reports of UK hospitalised children in which norovirus and adenovirus were detected in 15–16% and 14–15% of cases, respectively [168].

The high prevalence of adenovirus can be explained by the broad-range nature of the adenovirus PCR. Adenovirus is a DNA virus with several serotypes, each causing different clinical presentations. The serotypes most commonly associated with gastroenteritis are 40 and 41; however the real-time PCR assay detects, but does not distinguish between, all adenovirus serotypes. Immunocompromised patients can experience endogenous reactivation of persistent adenovirus infection [160]; adenovirus is detected in the stool of more than one third of paediatric HSCT patients

without associated clinical symptoms of acute intestinal infection [161]. Therefore it is possible that virus shed in stool is a manifestation of a focus of infection elsewhere in the body or reactivation of endogenous non-serotype 40/41 adenovirus. Consequently a proportion of the adenovirus infections detected in this study may not be causing gastroenteritis.

3.4.2 Norovirus in post-rotavirus vaccination era

Following the introduction of the rotavirus vaccine to the UK childhood vaccination programme in July 2013 the prevalence of rotavirus infections has reduced by 67% [164], which is reflected in the low prevalence of 2% reported in this study; an earlier study of hospitalised UK children reported a 31% rotavirus positive rate [168]. Whilst the rate of rotavirus positive patients is reduced compared to the pre-vaccination UK study [168] the rate of detection of norovirus, adenovirus, sapovirus and astrovirus is unchanged; consequently the overall positivity rate for gastrointestinal viruses is lower than previously reported; 23% in this study compared to 53% reported previously [168]. This suggests that other gastrointestinal viruses, including norovirus, have not increased in prevalence to fill the ecological niche left by diminishing rotavirus infections. Presumably this is because rotavirus and the other gastroenteric viruses, in addition to being antigenically diverse, are not competing to fill the same ecological niche therefore infection with one gastrointestinal virus does not preclude infections with another, as demonstrated here by the detection of mixed infections.

3.4.3 Viral titres

The retrospective nature of this study does not allow comparison of viral genome copy numbers between patients, as this is not routinely determined for norovirus in a diagnostic laboratory. However the linear relationship between PCR cycle threshold (Ct) values and genome copy numbers, in which a difference in approximately three cycles (3 Cts) is equivalent to a log10 difference in copy number, makes Ct values a good semi-quantitative indicator of viral titre and so allows us to compare viral titres based on Ct values. Furthermore the inherent variability of Ct values between PCR runs is small, thus Ct values are reproducible and we can report log differences with confidence.

Measuring viral loads in stool specimens is inherently difficult due to differences in consistency between specimens. The volume of stool added to an extraction is semi-standardised by adding a pea-size amount of stool to a fixed volume of lysis buffer, however the actual amount will vary; a pea-size amount of liquid stool is difficult to

measure. Differences in stool consistency make the input variable which may falsely indicate a higher or lower viral titre when comparing samples. Accuracy of viral load measurements can be improved by weighing the amount of stool that is extracted, and either standardising the input weight between specimens or expressing a viral load in copies/gram. However weighing stools may still produce variable results since the weight of liquid and solid stools may not be comparable.

The data presented in this chapter was exported, retrospectively, from the laboratory information system; since the results are from routine diagnostic specimen processing, stools were not weighed. Consequently viral titres should be interpreted with caution; the observations made, outlined below, should be verified in a prospective study that carefully controls the stool input prior to viral load measurement.

The majority (87%) of norovirus infections were genogroup II (GII), which had a significantly higher viral titre than GI infections. The higher titre seen in GII is not biased by infections in immunocompromised patients since there was no significant difference in PCR Ct values between immunocompromised and non-immunocompromised patients. The ratio of GI:GII infections is the same as that reported in smaller paediatric cohorts in the US [170] and in Venezuela [172]. The difference in viral titre between genogroups has also been observed in the US [170].

The average norovirus viral titre was higher in HAI compared to POA; the most parsimonious explanation for this is that HAIs are more likely to be detected nearer to the onset of symptoms, whereas POAs may be detected sometime after onset, thus a waning viral load is detected.

Despite the equal number of norovirus and adenovirus infections, norovirus infections are associated with a significantly greater virus titre with a difference in median PCR Ct value of 9 cycles. The difference in virus titre is more pronounced between norovirus and sapovirus, in which the observed difference in median Ct values is suggestive of a 10,000-fold difference in viral load. The lower limit of detection of the PCR assays used to detect each of the viruses is comparable (1–5 copies/µl for all except adenovirus 0.2 copies/µl), suggesting the observed difference in viral titres is not an artefact of assay sensitivity. However the observed Ct values are from the first positive stool sample from each infected patient; it is possible that if longitudinal sampling were performed one would observe a peak in viral titre in the other viruses later than is seen for norovirus.

3.4.4 Norovirus seasonality and genotypes

Unexpectedly, the overall prevalence of norovirus does not follow the characteristic seasonal trend seen in national outbreaks. Instead the number of infections per month strongly correlates with the number of hospital admissions, accounting for 50% of the variability in norovirus prevalence. It is possible that the increase in number of admissions is caused by heightened norovirus activity, rather than the other way around, however this is an unlikely scenario since GOSH has no A&E department therefore acute gastroenteritis is not the primary reason for admission. Moreover, the lack of wintertime seasonality does not reflect the national data from PHE suggesting admissions are not linked to heightened norovirus activity. Conversely, these results suggest that the prevalence of norovirus in a tertiary children's hospital is driven by traffic through the hospital, rather than seasonal outbreaks. The breadth of genotypes seen in this study, more commonly seen in community cohorts compared to hospitals, backs this hypothesis; patients presenting to primary healthcare facilities, such as GP practices, reportedly have a lower proportion of GII.4 infections; 54% compared to 91% of hospital infections are GII.4 [173]. The true distribution of norovirus genotypes in the community is not known since all genotyping studies to date are based on patients presenting to healthcare facilities thus introducing a presentation bias. There have been some comprehensive, unbiased, community studies on norovirus prevalence; namely Infectious Intestinal Disease (IID) [174], the Second Study of IID in the UK (IID2) [113] and the Malnutrition and the Consequences for Child Health and Development (MAL-ED) study [175]. IID was a population based community cohort study in England in 1995, with 9776 randomly selected community participants from 1995 to 1996; IID2 was a UK multicentre longitudinal community cohort study conducted in 2008–2009, with 6836 participants; MAL-ED is a multi-country birth cohort of 1457 children across eight countries (South Africa, Tanzania, India, Pakistan, Bangladesh, Nepal, Brazil and Peru) describing norovirus epidemiology during the first two years of life, with specimens collected from 2009 to 2014. However, to date, none of the norovirus infections from these studies have been genotyped. Genotyping of norovirus infections in these unbiased community cohorts is needed in order to determine whether infections caused by a breadth of genotypes are a true reflection of the community.

In this study cohort just over half (52%) of norovirus infections were identified as GII.4, which is the dominant norovirus genotype worldwide. A quarter (26%) of infections were caused by GII.3, which has previously been described in varying proportions in UK cohorts, from 0–35% early in the norovirus season (September/October) to 0–10%

late in the season (March/April) [129]. The high prevalence of GII.3 in our paediatric population is different to adult cohorts, in which outbreaks are largely dominated by GII.4 [173] but in line with Sukhrie *et al.*'s finding of an association between GII.3 and paediatric wards [140]. A peak of GI.3 from September to December 2014 mirrors a similar peak in GI.3 in national outbreaks reported to the PHE, suggesting that the virus circulating in the hospital is to some extent reflective of the national landscape. However this is not the case for GII.3 which is only reported sporadically in national outbreaks. The data in this study supports the notion that GII.3 is more frequently associated with children [140]; however the reason for this is unknown. Speculatively, the reason could be immunity to GII.3 in the adult population following childhood infection, lower transmissibility compared to GII.4 resulting in fewer associated outbreaks and reporting bias or differences in receptor binding between children and adults. The latter suggestion is of particular interest since recent success in culturing human norovirus *in vitro* demonstrated that GII.3 (and other non-GII.4 genotypes) require additional co-factors for cell entry that are not required by GII.4 viruses [34]. This warrants further investigation but is beyond the scope of this study.

3.4.5 Norovirus in clinical specialties

All patients in this study have been categorised by clinical specialty; this was based on the ward to which they were admitted at the time of specimen collection. It is clear that immunocompromised patients are over-represented among patients with norovirus infection; 19% of immunocompromised patients were found to be norovirus positive, compared to just 5% and 7% of surgical and medical patients, respectively. Previous studies in smaller cohorts of 47, 61,78 and 116 immunocompromised patients have reported prevalence of norovirus as 23%, 21%, 17% and 22%, respectively [54, 116, 117, 176] which suggests the categorisation of patients into clinical specialties in this study is reliable and that the larger cohort of 270 immunocompromised patients in this study corroborates earlier findings.

There is no difference in age between immunocompromised and medical/surgical patients tested for viral pathogens in this study (P=0.972, 2-sample T-test). The duration of hospital stay for each patient is not known, however there is a sampling bias between immunocompromised and medical/surgical patients, with more stool specimens tested for norovirus from immunocompromised patients compared to medical/surgical patients (mean 5 and 3 specimens per patient, respectively, P<0.001, 2-sample T-Test), which may falsely indicate a higher prevalence of norovirus in immunocompromised patients. Since this is an observational study using routine

clinical data, a prospective study without sampling bias is required to confirm the observed results.

3.4.6 Chronic infections

Norovirus caused the highest rate of chronic infections (26% PCR positive >1 month). Nonetheless all other viruses tested in this study were also shown to cause chronic infections; 15–19% in sapovirus, rotavirus and adenovirus infections and 6% in astrovirus infections. The vast majority of chronic infections were in immunocompromised patients.

Chronic norovirus infections in immunocompromised patients is a recognised cause of morbidity, in whom a bi-phasic illness develops [177] with an initial acute phase followed by a second chronic phase with viral shedding and diarrhoea lasting weeks to years. The consequence of chronic norovirus infection can be dehydration, malnutrition, dysfunction of intestinal barrier [80], dramatic weight loss [81], a requirement for nutritional support [54] and in extreme cases death [81, 82]. Immunocompromised patients in this study do not show a higher norovirus viral titre or difference in genotypes; suggesting the higher chronicity in immunocompromised patients is host, not virus, mediated.

3.4.7 Norovirus in B cell deficient patients

This study shows that susceptibility to norovirus infection between patients with and without B cells does not differ.

The finding that the average viral titre (based on Ct values) in B cell deficient patients is significantly lower, albeit in a small cohort, is in agreement with Jones *et al.* who showed that B cell deficient mice had decreased viral titres compared to wild-type mice [24]. Nevertheless our finding of a mean norovirus Ct value of 22 in B cell deficient SCID patients indicates a very high viral titre, suggestive of millions of viral genome copies per ml. Therefore extensive viral replication must be occurring in host cells other than B lymphocytes. The finding presented here that B cells cannot be the primary site of norovirus replication *in vivo* has since been corroborated by Ettayebi *et al.* [34], who demonstrated *in vitro* that in human intestinal enteroids (HIEs), which contain multiple intestinal epithelial cell types, norovirus cell entry and replication is only supported in enterocytes. This, together with earlier histological data in intestinal biopsies from chronically infected immunosuppressed patients showing viral proteins in enterocytes [33], suggests that the primary site of human norovirus replication *in vivo* is enterocytes.

An alternative explanation of the results presented here is that the detection of norovirus in these patients represents repeated re-infection, rather than a single infection with on-going replication. However all patients were screened for the presence of norovirus on a weekly basis (with exception of Px 12 who was screened monthly). Residual specimens were not available to confirm on-going infection based on genome sequencing, however all samples from norovirus positive-patients were consecutively positive for norovirus with no negative samples in between; given the frequency of sampling it is unlikely that these infections represent repeated re-infections.

This unique B cell deficient patient cohort at GOSH provides data in humans to support the *in vitro* and animal findings reported by Jones *et al*.; the difference in viral titres suggests a potential role for B cells in norovirus replication. However the occurrence of norovirus infection in B cell deficient patients suggests there must be other primary sites for norovirus replication. This finding is supported by histopathological analysis of intestinal biopsies by Karandikar *et al.* (2016), which did not detect norovirus proteins in B cells but did detect viral proteins in enterocytes [33], and by Ettayebi *et al.* (2016) who successfully cultivated human norovirus in enterocytes generated from intestinal stem cells [34]. Together with norovirus infection of B cell deficient patients presented in this chapter, this suggests B cells are not the primary site of infection *in vivo*.

3.4.8 Conclusions

This chapter reports the prevalence of gastroenteric viruses in a large observational cohort of paediatric patients in a UK hospital following the implementation of routine rotavirus vaccination. There is no change in the prevalence of other viruses compared to pre-vaccination studies. All viruses, but most commonly norovirus, are shown to establish chronicity, primarily in immunocompromised patients. New norovirus infections are not driven by seasonal trends, which may be specific to this population but has been reported in community cohorts [119]. The high proportion of non-GII.4 infections in children may have implications for vaccine development.

The prevalence of norovirus infection does not differ between patients with and without B cells, signifying B cells are not the primary site of *in vivo* norovirus replication. Nonetheless a difference in viral titre between the two patient groups suggests B cells may play a secondary role in norovirus replication.

3.5 ACKNOWLEDGEMENTS

The norovirus real-time PCR and genotyping assays were developed by me for this thesis, including data on assay performance. Great Ormond Street Hospital Virology department provided the diagnostic service which generated the diagnostic real-time PCR results presented in this study. The norovirus genotyping was performed by me for the first 3 months and by Divya Shah, under my supervision, for the remaining 9 months.

Figure 3(b) was reproduced, with permission, from the Public Health England (PHE) norovirus surveillance report. Rotavirus vaccine/wild-type typing was undertaken by the PHE Virus Reference Department.

CHAPTER 4

NOROVIRUS WHOLE GENOME SEQUENCING BY SURESELECT TARGET ENRICHMENT

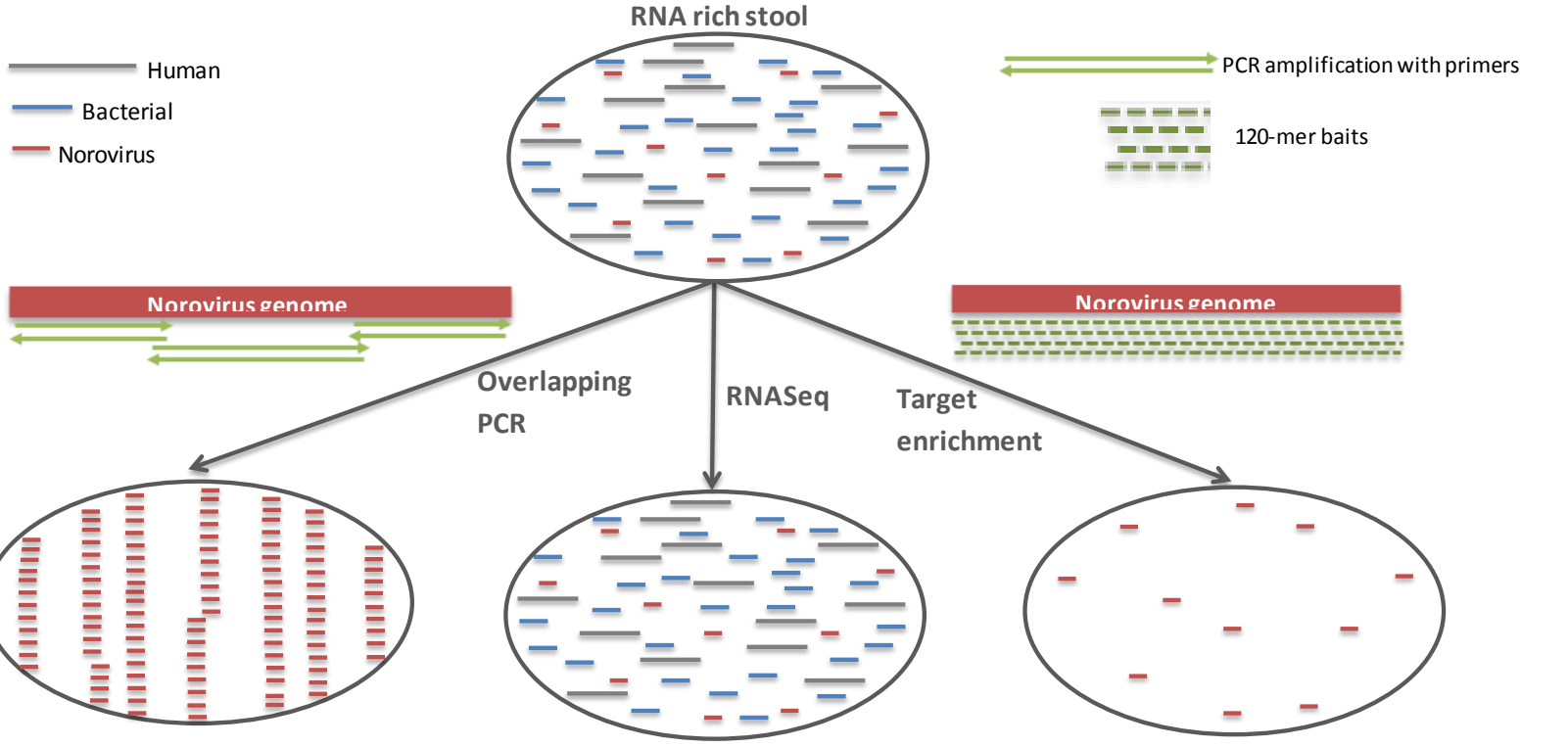
4.1 INTRODUCTION

Comparison of norovirus genetic sequences has the potential to allow linking of previously unrecognised transmission events or exclusion of cases from an outbreak. Traditionally, norovirus genotyping has involved polymerase chain reaction (PCR) amplification and capillary sequencing of partial regions of the polymerase and capsid sequences, followed by additional sequencing of the P2 region for outbreak investigations. This is a labour intensive process requiring several rounds of PCR and sequencing, each requiring genogroup or genotype specific primers and only yields partial genome sequences at the end. Whilst the P2 domain can identify linked outbreak events with 64–73% specificity (assuming bootstrap support >70 or <70, respectively), the full capsid sequence can identify linked outbreak events with 100% specificity [145] and thus is more informative.

Whole genome sequencing simplifies investigation of norovirus molecular epidemiology by generating all the regions of interest in one step, thus allowing identification of the genotype, variant type and full capsid sequence; negating the need for sequential PCR and sequencing reactions. However, unlike bacteria, which can be isolated in pure culture, norovirus culture is difficult [25, 34]. Moreover, as norovirus replicates within the host cell, viral nucleic extracts from culture would be contaminated by host DNA, and if obtained from clinical specimens, by DNA and RNA from enteric bacteria and host cells.

Two principle methods have been applied to sequence norovirus full genomes from clinical specimens to date; sequencing of overlapping polymerase chain reaction (PCR) fragments and direct sequencing of total RNA. A third method, target enrichment, is the subject of this chapter. The overlapping PCR generates pure viral template, which results in sequencing data that is all on-target, but requires multiple PCR amplifications. Total RNA sequencing necessitates great depth of sequencing to generate the target norovirus genome. All three methods are summarised in **Figure 4.1**. The latter two methods, total RNA sequencing and target enrichment, have been achieved with the advent of next-generation deep sequencing.

Figure 4.1 Schematic overview of the three principle methods used to generate norovirus full genome sequences with summary of advantages and disadvantages for each method



Pros	Cons
Generates high yields of target DNA	Sequence heterogeneity problematic for primer design therefore primers are often genotype specific. High failure rate, especially with non-target genotypes

Pros	Cons
Direct sequencing of total RNA; no need for primer design No prior knowledge of sequences required Not genotype specific	Majority of data is redundant with low proportion of reads generated corresponding to norovirus, resulting in low read depth and limited scope for variant analysis. Limited success with low titre samples

Pros	Cons
Bait design accounts for sequence heterogeneity Not genotype specific Good read depth, sufficient for variant analysis Successful with low titre samples	Bait design relies on availability of sequences; novel genotypes that are highly divergent from known genotypes may have limited success

4.1.1 Overlapping PCR amplicon sequencing

The advantage of PCR for sequencing from clinical specimens is that over the course of 30 to 40 amplification cycles the target sequence, in this case norovirus, is amplified exponentially generating a high yield of DNA for sequencing. The non-target sequences become insignificant. The generated amplicons can then be sequenced either by capillary sequencing or with a next-generation sequencing (NGS) platform.

The norovirus genome is 7.5kb. Since this is beyond the upper size limit of a PCR reaction, the genome must be amplified in several overlapping fragments, referred to as tiled PCR. PCR amplification works using specific primers that are designed to be complementary to a conserved region of the target sequence and will not amplify sequences that do not match both the forward and reverse primers. This is problematic for norovirus due to the sequence heterogeneity both within and between genotypes.

Smaller fragments (<1kb) amplify with greater efficiency than larger fragments, and thus was the approach we used to produce pilot data preceding this thesis to successfully amplify norovirus full genomes using 22 overlapping PCR fragments [148]. However not all of the primer sites are conserved between genotypes therefore this approach could only be used to amplify GII.4 sequences. Larger fragments (2–3 kb), whilst more difficult to amplify, require fewer primers thus are more likely to be successful across genotypes. Nonetheless due to sequence heterogeneity both between and within genotypes this approach retains a limited success rate, as demonstrated by Cotten *et al.* [149] who successfully amplified full genome sequences from 83%, 88%, 92% and 100% of GII.13, GII.6, GII.4 and GII.9 samples but only 20%, 40%, 50% and 77% of GI, GII.2, GII.12 and GII.3 and 0% of GII.7 samples.

Norovirus whole genome sequencing from a single 7.5 kb amplicon has been described and used to generate 25 full genome sequences [146]. Although the authors do no report the success rate using this approach, it is generally very difficult to amplify fragments of such a size. The long dwell times of the PCR cycling often result in non-specific amplification and the technique relies on purifying intact full length RNA genomes; fragmented genomes are not amplified.

4.1.2 Direct Sequencing of total RNA (RNASeq)

The most straight forward method to sequence norovirus genomes direct from stool is whole transcriptome sequencing, or RNASeq. This involves purifying the total RNA content of a stool specimen and preparing a sequencing library with all of the RNA,

which is then sequenced on an NGS platform. To avoid unnecessary sequencing of host ribosomal RNA, library preparations can either remove ribosomal RNA (ribodepletion) or select polyA-tailed mRNA (the transcriptome) which will include polyA-tailed norovirus genomes.

The advantage of RNASeq is that there is no requirement for PCR primers therefore it is completely unbiased. Although norovirus whole genomes generated by RNASeq reported to date are predominantly GII.4, it is theoretically possible to sequence all genotypes with equal success.

The biggest limitation of RNASeq is the same as its very reason for success. The unbiased nature of RNASeq means that all RNA present in the sample, which will include host and bacterial RNA transcripts as well as norovirus RNA, will be sequenced. The proportion of norovirus to other RNA can be very low, therefore large sequencing capacity is required to capture the norovirus sequences in each sample. Since sequencing platforms generate an approximately fixed amount of data, which is shared among the number of samples sequenced on one run, to generate the read depth required to recover norovirus genomes the number of samples sequenced per run is very limited. For instance to generate norovirus full genomes Illumina's bench-top sequencing platform MiSeq, which generates approximately 15 Gb of data and 25 million reads, can only accommodate 6 stool samples per run [178].

The data generated by RNASeq is sufficient to generate almost complete norovirus genome sequences; 40–99% of reported samples achieved >90% genome coverage. However due to the low proportion of reads that belong to norovirus, median 2–3% across all reported samples, the average read depth per sample sequenced on a MiSeq is only 9–259 [147, 178, 179]. Whilst this read depth is, in most cases, sufficient for a reliable consensus sequence, it does not allow variant analysis. Sequencing on a higher-throughput platform, such as the Illumina HiSeq which generates up to 750 Gb and 2.5 billion reads, allows up to 96 samples to be sequenced on a single run. Whilst the proportion of reads generated that corresponds to norovirus is still only 3%, the increased sequencing output allows greater read depth per sample, in some cases up to 1000-fold. Nevertheless the median reported mean depth remains approximately 100-fold [178].

Unbiased sequencing of non-target RNA results in low read depths but also has a financial implication; for instance the reagents for a single MiSeq or HiSeq sequencing run (not including library preparation) cost approximately £600 or £4000, respectively;

when shared between few samples this does not make whole genome sequencing cost-effective.

4.1.3 Target enrichment

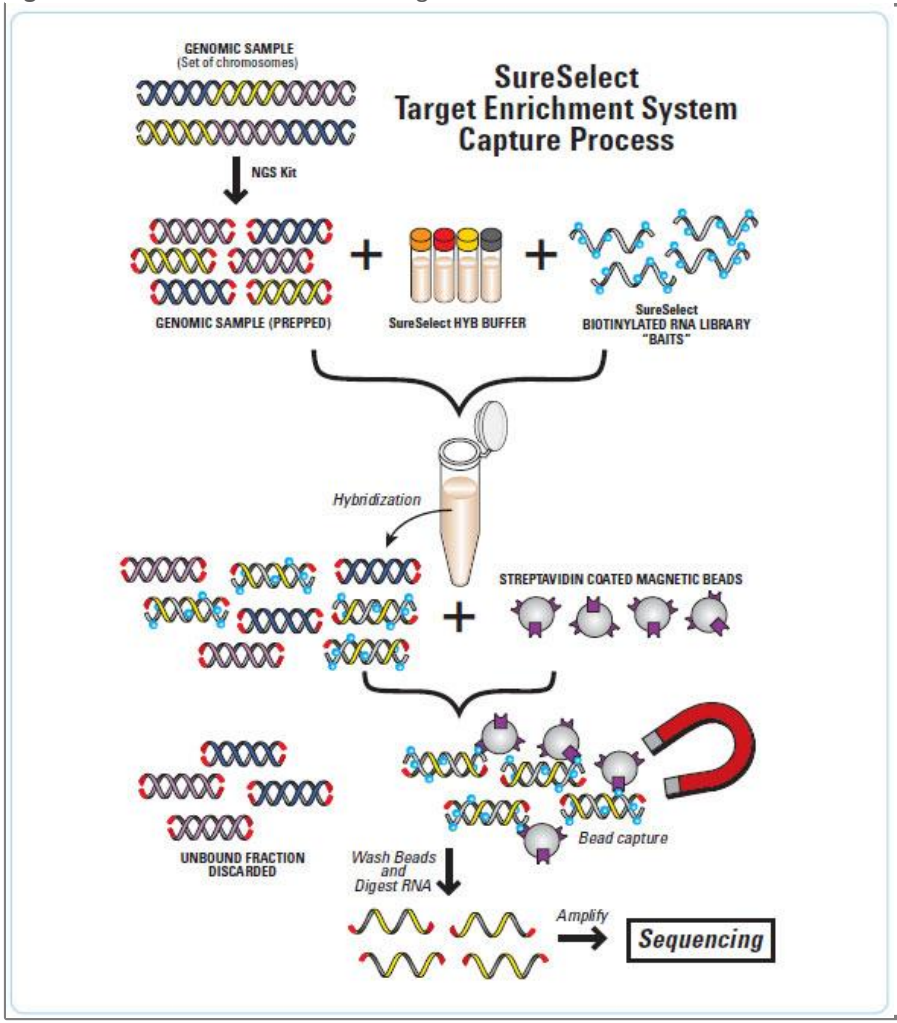
The proportion of norovirus RNA in stool samples can be very low, resulting in low read depths when directly sequencing total RNA. The purpose of target enrichment is to purify the RNA of interest in the specimen, in this case norovirus, and discard the rest. Consequently the specimen is enriched for the genomes of interest and sequencing isn't wasted on non-target genomes, such as host or bacteria.

Enrichment makes use of a panel of custom-designed 120-mer RNA baits which are complementary to all publicly available target genome sequences, with multiple baits targeting each position of the genome. Baits are tiled across the genome so that every position in the genome has a complimentary bait. cDNA synthesised from total RNA is incubated with the biotinilated target-specific RNA baits, which will hybridise to target genomes present in the sample. Streptavidin coated magnetic beads will in turn bind to the hybridised baits, which can then be separated from all non-target cDNA using magnetic bead capture. The RNA baits are then digested, leaving an enriched library of target cDNA which can be sequenced on an NGS platform (**Figure 4.2**). Unlike PCR which introduces a bias for A to G and T to C substitutions, it has been demonstrated that target enrichment does not introduce nucleotide substitution bias [180]. The great advantage of this technique is that, unlike PCR which uses a single primer at each site, multiple baits are designed to cover each position, thus accounting for sequence variation between norovirus genomes. This has the potential for un-biased sequencing across genotypes in a single reaction, without prior knowledge of the genotype present in the specimen. Nonetheless, as with overlapping PCR, target enrichment does require pre-existing knowledge of norovirus sequence diversity for bait design. Should a novel genotype emerge which is not included in bait design, or an existing genotype evolve to the point where it is no longer enriched by existing baits, this will result in failure of target enrichment.

SureSelect target enrichment was initially developed for human exome sequencing [181]. However it has since been translated to full genome sequencing of viruses or difficult to culture bacteria; the objective is to enrich for whole genomes of a specific pathogen in a background of other, non-target, genomes. Pathogens that have successfully been sequenced to date using target enrichment are varicella zoster virus [180], hepatitis C virus [182], human herpes virus 6 [183], human herpes virus 7 [184],

Mycobacterium tuberculosis [185] and *Chlamydia trachomatis* [186]. This is its first application for sequencing norovirus genomes.

Figure 4.2 Schematic of SureSelect target enrichment workflow for norovirus whole genome sequencing. Image reproduced, with permission, from Agilent Technologies



cDNA synthesised from stool total RNA

cDNA hybridised to 120 nt biotin-labelled norovirus-specific RNA baits designed to cover entire genome. Multiple baits at each position in the genome accommodate sequence heterogeneity between genomes

Magnetic beads bind biotin-labelled DNA-RNA fragments

Target cDNA pulled out using magnet

Norovirus cDNA is enriched and ready for deep sequencing

4.2 MATERIALS AND METHODS

4.2.1 Samples

507 norovirus positive stool samples from 382 patients in four UK healthcare institutions were processed for whole genome sequencing. Samples included genotypes GI.1, GI.2, GI.3, GI.6, GI.7, GII.1, GII.2, GII.3, GII.4, GII.5, GII.6, GII.7, GII.13, GII.14 and GII.17, as detailed in **Table 4.1** on Page 121. The presence of norovirus was verified in all samples using a multiplex norovirus GI and GII-specific one-step reverse-transcription real-time PCR (RT-qPCR); the primer and probe sequences and cycling conditions are described in Chapter 2. For 78/507 samples provided by one of the centres, the presence of norovirus RNA was not verified in the re-extracted residual specimen; for these samples the RT-qPCR Ct value corresponds to the original extract used as part of diagnostic service. The RT-qPCR cycle threshold (Ct) value is used in this study as a semi-quantitative indicator of viral titre.

All specimens were residual diagnostic specimens obtained from patients with confirmed norovirus infections. Specimens were submitted to the UCL Infection DNA Bank for use in this study. All samples were supplied to the study in an anonymised form; the use of these specimens for research was approved by the NRES Committee London – Fulham (REC reference: 12/LO/1089). All stool samples were stored at −80°C in between diagnostic testing and RNA extraction for full genome sequencing.

164 stool samples were genotyped using capsid PCR and Sanger sequencing in parallel to SureSelect target enrichment whole genome sequencing. PCR primer sequences and cycling conditions for capsid genotyping are described in Chapter 2. Briefly, GI or GII-specific primers were used to amplify a 597 or 468 nt region of the norovirus capsid shell domain, respectively; amplicons were capillary sequenced in the forward and reverse direction. Generated sequences were submitted to the Norovirus genotyping tool to identify the capsid genotype [141].

4.2.2 Specimen processing and sequencing

The details for RNA extraction, cDNA synthesis, SureSelect bait design, target enrichment, sequencing library preparation and Illumina sequencing are described in detail in Chapter 2.

4.2.3 Negative controls

All RNA extraction batches included a negative extract control, consisting of sterile Qiagen Buffer ASL extracted with the Qiagen EZ1 virus mini kit alongside stool

samples. All negative extracts were tested by norovirus-specific RT-qPCR to verify the absence of contaminating RNA.

To determine the level of contaminating norovirus RNA in the sequencing pipeline, two negative extracts were processed for sequencing.

4.2.4 Sequence assembly

All assemblies were done in CLC genomics workbench v8 using a *de novo* assembly pipeline, as described in detail in Chapter 2.

4.2.5 Simulated mixed infection

To assess whether a reliable consensus sequence can be generated from a mixed infection, the reads generated from two single infections (one GII.3 and one GII.4) were merged into a single assembly pipeline. The consensus sequences generated from the single infection (original) and the mixed (simulated) infection were aligned to identify the number of differences between the two consensus sequences.

4.2.6 Sequencing outcomes and parameters

For each sequence generated, the following outcomes were noted. First, the proportion of sequencing reads that map to a norovirus reference during the assembly filtering step referred to as the percent on-target reads (% OTR). Second, the proportion of the norovirus genome which the generated sequence covers referred to as the % genome coverage; this is based on the expected genome size for the relevant genotype (e.g. 7560 nt for GII.4 genomes). Third, the average number of reads mapping at each position across the genome, referred to as the mean read depth.

Parameters were set to determine sequence success. The cut-off for sequencing success was defined as >90% coverage of the full norovirus genome with >100-fold mean read depth. These parameters were chosen based on the requirement to generate complete genome sequences (>90% genome coverage) with robust consensus sequences (>100-fold read depth). A 100-fold minimum read depth was chosen because, although in bacterial genomes a 20-fold read depth is considered sufficient to overcome sequencing errors [187], since viral genomes exist as a quasispecies [51] a higher read depth was chosen to maximise accuracy at polymorphic sites. A minimum read depth of 100-fold has previously been used as the minimum read depth for identification of polymorphic loci in DNA viruses [180]. Samples that met only one of these criterions were categorised as “sub-optimal”, and those which did not meet either criteria were considered a “fail”.

4.2.7 Statistical analysis

All statistical analysis was performed in SPSS v23 as described in Chapter 2.

4.2.8 Evaluation of turn-around times

To measure turn-around times for sequencing in a clinical setting, for five weeks (January–February 2016) all of the first-time norovirus positive stool specimens at GOSH were sequenced prospectively, in weekly batches. Each batch consisted of 1–12 new norovirus specimens; to remain cost-effective batches were made up to 12 using archived study samples. The 12 norovirus samples were sequenced in multiplex alongside other pathogen samples, in order to achieve 48–96 samples per MiSeq run.

4.3 RESULTS

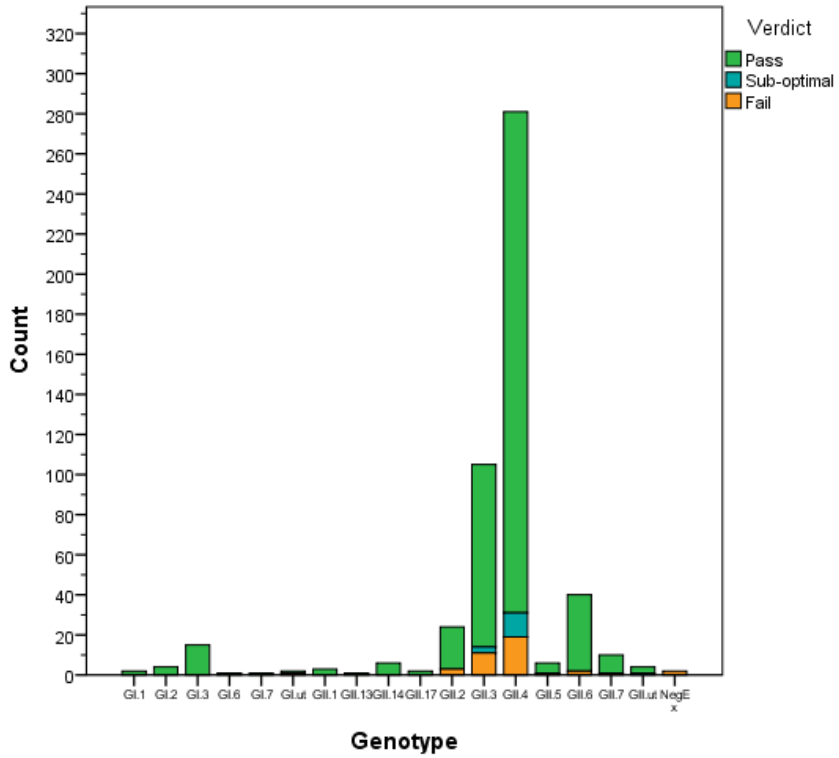
4.3.1 Overall sequencing outcomes

A median of 81% of the total sequencing reads generated for each sample mapped to the norovirus genome, referred to as the % on-target-reads (% OTR). On average, 100% of the full genome was covered (% genome coverage) with median read depth of 12,227-fold (**Table 4.1**).

Of 507 samples across all sampled genotypes, 453 (89%) passed; i.e. had >90% genome coverage and >100-fold read depth. 93% of samples had a genome coverage of >90% at any depth. (**Table 4.1**, **Figure 4.3**).

Figure 4.3 (a) Number of samples sequenced according to norovirus genotype, classified by sequencing outcome; **(b)** Sequencing outcome proportions for each norovirus genotype. Pass, >90% genome coverage and >100-fold read depth; Sub-optimal, >90% genome coverage or >100-fold read depth; Fail, <90% genome coverage and <100-fold read. Genotype refers to capsid genotype only

(a)



(b)

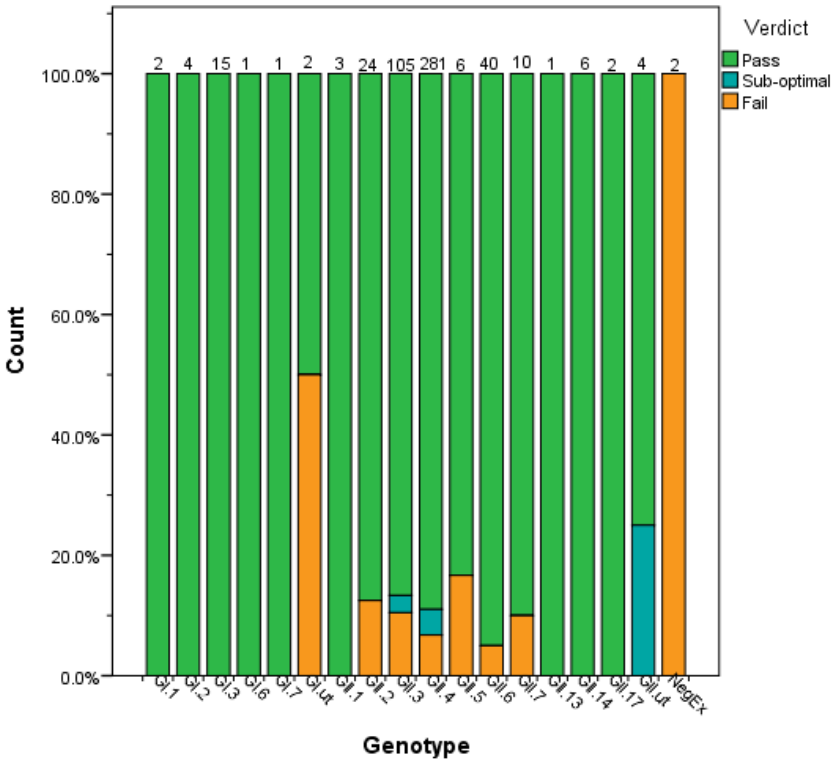


Table 4.1 Metrics of norovirus whole genome sequencing for all samples (TOTAL) and for each genotype

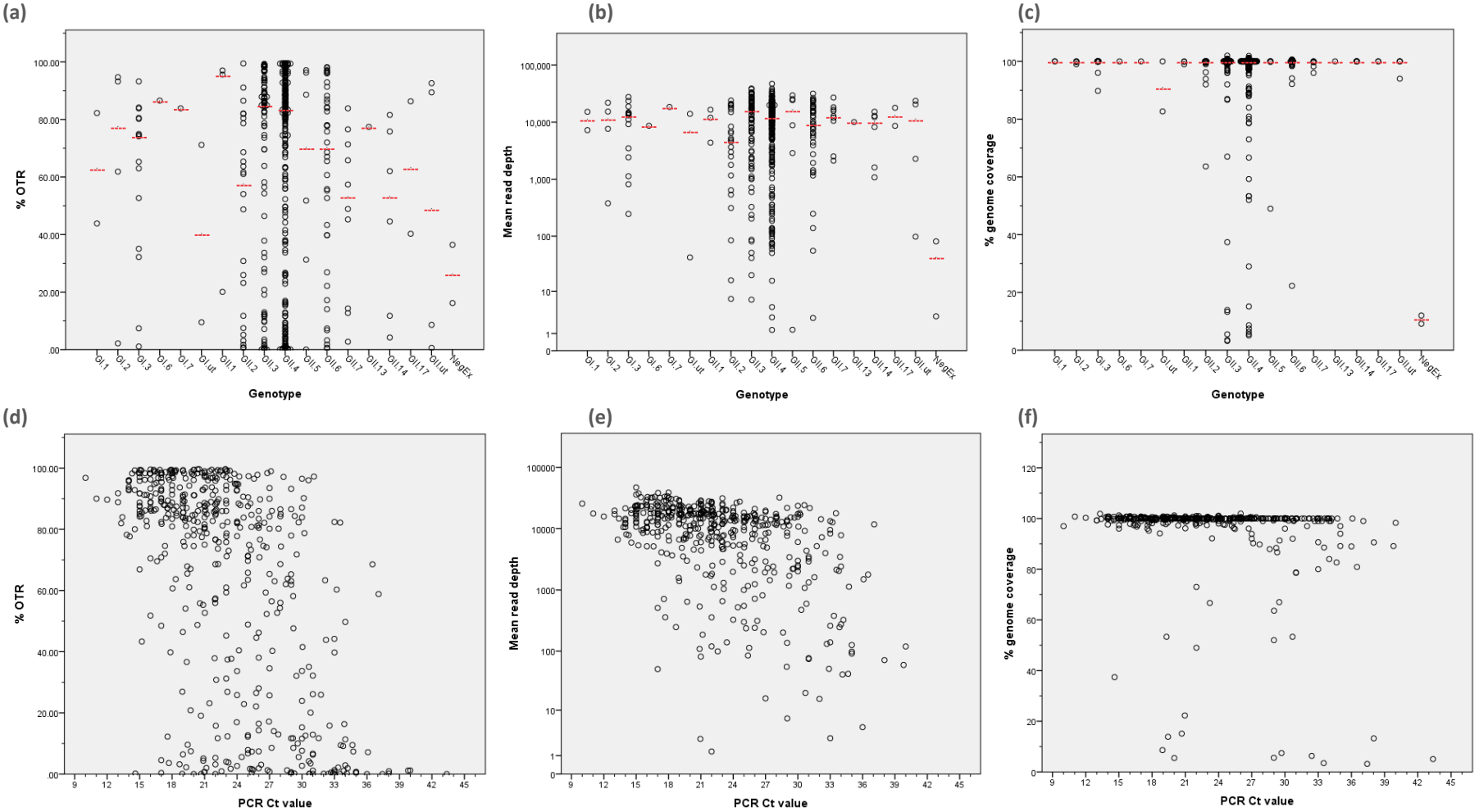
	Number of samples sequenced	Number samples Pass (%)	Number samples Sub-optimal (%)	Number samples Fail (%)	Median % OTR (min–max)	Median read depth (min–max)	Median % genome coverage (min–max)	Median Ct values (min–max)
GI.1	2	2 (100%)	0 (0%)	0 (0%)	63.05 (43.85-82.25)	11,194 (7,239–15,149)	100 (100–100)	31 (30–32)
GI.2	4	4 (100%)	0 (0%)	0 (0%)	77.60 (2.17-94.70)	11,464 (379–21,843)	100 (99–100)	29 (24–33)
GI.3	15	15 (100%)	0 (0%)	0 (0%)	74.13 (1.08-93.25)	13,157 (246–27,569)	100 (90–100)	27 (17–35)
GI.6	1	1 (100%)	0 (0%)	0 (0%)	86.56 (n/a)	8,642 (n/a)	100 (n/a)	29 (n/a)
GI.7	1	1 (100%)	0 (0%)	0 (0%)	83.88 (n/a)	18,414 (n/a)	100 (n/a)	21 (n/a)
GI.ut	2	1 (50%)	0 (0%)	1 (50%)	40.34 (9.50-71.18)	7,000 (42–13,957)	91 (83–100)	29 (23–35)
GII.1	3	3 (100%)	0 (0%)	0 (0%)	95.61 (20.06-97.04)	11,990 (4,365–16,506)	100 (99–100)	15 (14–31)
GII.13	1	1 (100%)	0 (0%)	0 (0%)	77.44 (n/a)	10,043 (n/a)	100 (n/a)	21 (n/a)
GII.14	6	6 (100%)	0 (0%)	0 (0%)	53.31 (4.20-81.60)	10,238 (1,081–15,215)	100 (100–100)	27 (21–32)
GII.17	2	2 (100%)	0 (0%)	0 (0%)	63.30 (40.27-86.33)	13,204 (8,598–17,811)	100 (100–100)	24 (21–27)
GII.2	24	21 (88%)	0 (0%)	3 (12.5%)	57.60 (0.60-99.47)	4,717 (7–23,889)	100 (64–100)	24 (18–32)
GII.3	105	91 (87%)	3 (2.9%)	11 (10.5%)	85.00 (0.02-99.36)	16,034 (7–38,843)	100 (3–100)	21 (10–38)
GII.4	281	250 (89%)	12 (4.3%)	19 (6.8%)	83.75 (0.02-99.63)	12,465 (1–46,996)	100 (5–100)	22 (12–43)
GII.5	6	5 (83%)	0 (0%)	1 (16.7%)	70.21 (0.04-97.13)	16,468 (1–29,488)	100 (49–100)	19 (16–23)
GII.6	40	38 (95%)	0 (0%)	2 (5%)	70.32 (0.45-98.23)	9,356 (3–31,643)	100 (22–100)	21 (13–33)
GII.7	10	9 (90%)	0 (0%)	1 (10%)	53.14 (2.72-83.88)	12,779 (2,106–26,914)	100 (96–100)	25 (22–30)
GII.ut	4	3 (75%)	1 (25%)	0 (0%)	49.02 (0.59-92.61)	11,356 (98–23,588)	100 (94–100)	25 (19–35)
NegEx	2	0 (0%)	0 (0%)	2 (100%)	26.30 (16.18-36.42)	42 (3–81)	11 (9–12)	Not detected
TOTAL	**509**	**453 (89%)**	**16 (3%)**	**40 (8%)**	**81.22 (0.02–99.63)**	**12,227 (1–46,996)**	**100 (3–100)**	**22 (10–43)**
Total excl. Run 30 & 31	**413**	**381 (92%)**	**16 (4%)**	**16 (4%)**	**84.45 (0.02–99.63)**	**14,341 (1–46,996)**	**100 (13–100)**	**22 (10–40)**

Pass, >90% genome coverage and >100-fold read depth; Sub-optimal, >90% genome coverage or >100-fold read depth; Fail, <90% genome coverage and <100-fold read depth; n/a, range not applicable due to single sample; % OTR, percent on target reads; Ct, real-time PCR cycle threshold; GI.ut, genogroup I untypable; GII.ut, genogroup II untypable; NegEx, negative control

There was no significant difference in % OTR (P = 0.127), mean read depth (P = 0.398) or % genome coverage (P = 0.203) between norovirus genotypes (**Figure 4.4 (a–c)**).

A significant correlation was found between % OTR and read depth (R = 0.757, P <0.001, **Figure 4.5**) and between PCR Ct value and (i) % OTR (R = −0.536, P <0.001), (ii) read depth (R = −0.468, P <0.001) and (iii) % genome coverage (R = −0.223, P <0.001) (**Figure 4.4 (d–f)**). It follows that there is a significant difference in PCR Ct value between samples that passed compared to those that were sub-optimal (P <0.001) or failed (P <0.001) with median Ct values of 22, 32 and 32, respectively (**Figure 4.6**). There is an inverse relationship between Ct value and viral load [188]; thus samples with a smaller Ct value (higher viral titre) resulted in higher %OTR, read depth and genome coverage.

Figure 4.4 Norovirus full genome sequencing outcome metrics according to **(a–c)** norovirus genotype and **(d–f)** RT-qPCR Ct value. Red lines indicate median value



% OTR, % on-target-reads; Ct, cycle threshold

Figure 4.5 Correlation of read depth and % OTR across all samples (n = 507). R = 0.757, P <0.001

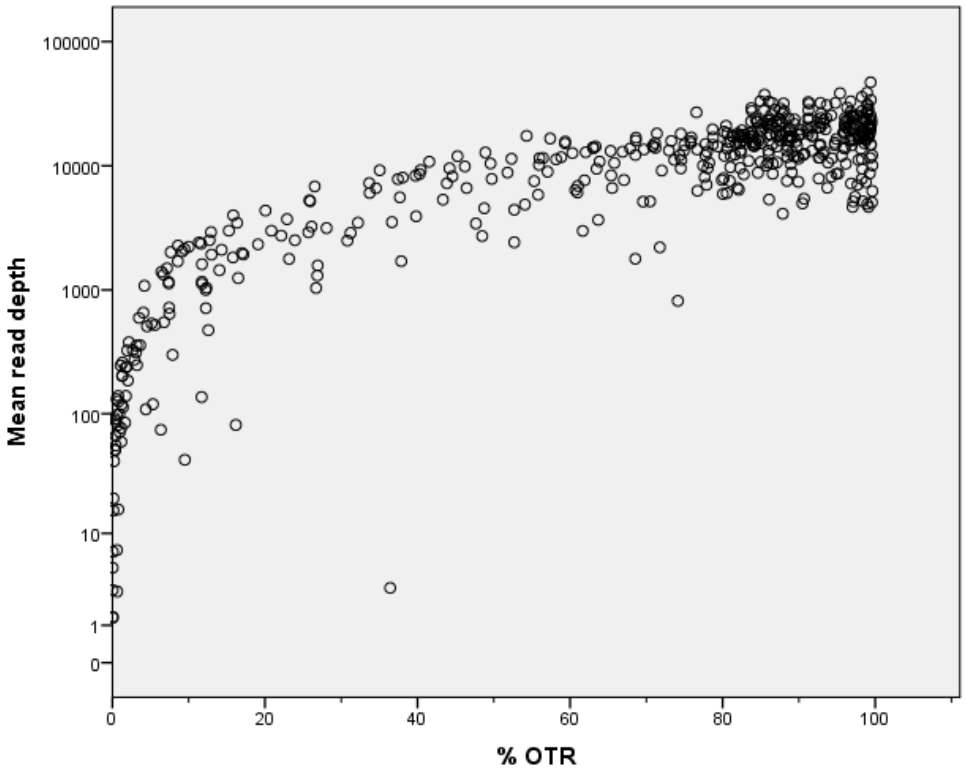
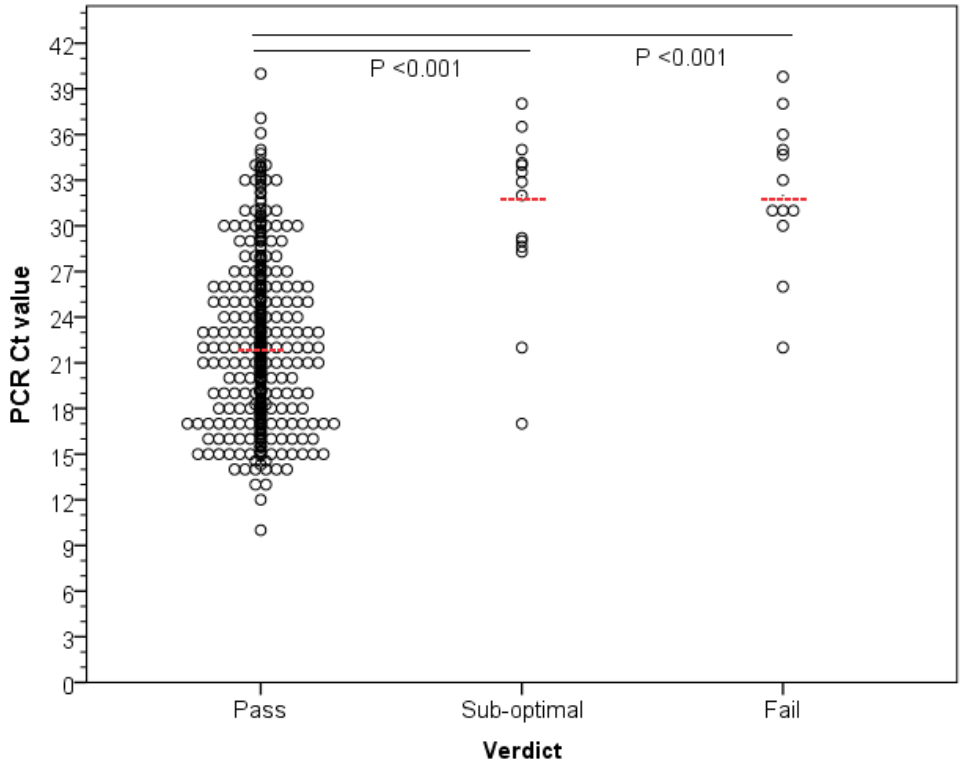


Figure 4.6 RT-qPCR Ct value of all samples, excluding Run 30 and 31, (n = 413) sequenced by SureSelect. Pass, >90% genome coverage and >100-fold read depth; Sub-optimal, >90% genome coverage or >100-fold read depth; Fail, <90% genome coverage and <100-fold read depth

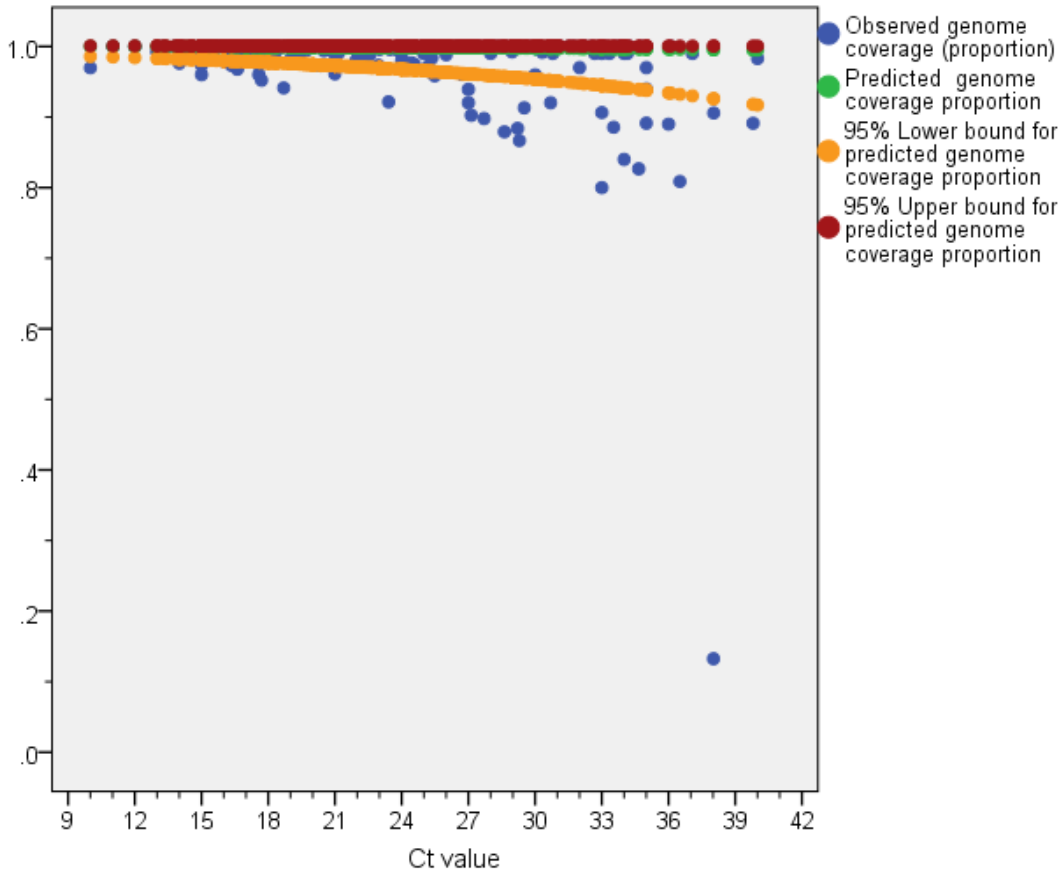


4.3.2 Predicted genome coverage

The estimated linear regression model is $y = 7.432 - 0.059x$ where the dependent variable *y* is the logit of transformed genome coverage proportion and the independent variable *x* is the PCR Ct value (n = 477, R^2 = 0.058, P <0.001).

Prediction intervals generated using the linear regression model predict that stool samples with a norovirus RT-qPCR Ct value <40 will generate 92–100% of the full genome sequence, with 95% certainty (**Figure 4.7**).

Figure 4.7 Observed and predicted % genome coverage values with 95% prediction intervals, excluding outliers detailed in section **2.20.2**). Fitted linear regression model: y=7.432-0.059x where the dependent variable y is the logit transformed genome coverage proportion

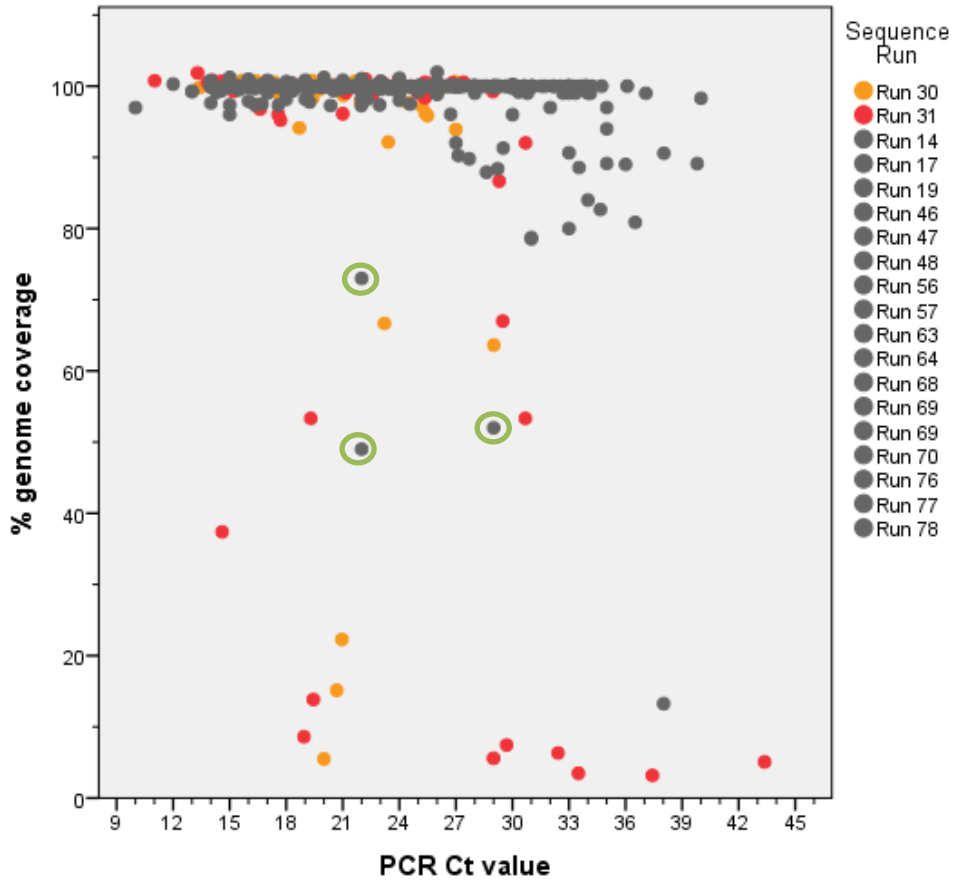


4.3.3 Failed samples

There were 22 (22/507) samples that achieved less than 80% genome coverage; including 15 with Ct values below 30 (suggesting a high viral titre) (**Figure 4.4 (f)**). Since all other samples (485/507) achieved greater than 80% genome coverage, including 414 samples with Ct below 30, these were considered to be outliers. These outliers are dominated by samples from two sequencing runs (#30 and 31; **Figure 4.8**), accounting for 18/22 of the outliers. Runs 30 and 31 had processing problems during

cDNA preparation. Specifically, cDNA for samples in Runs 30 and 31 was prepared in a 96-well plate; during RNA concentration in the vacuum centrifuge the rate of evaporation across the plate was unequal, resulting in slightly varied volumes for the next step of cDNA synthesis.

Figure 4.8 % genome coverage and Ct value, with sequencing runs 30 and 31 highlighted. Green circles highlight outliers that cannot be explained by sequencing runs 30 and 31



Three samples (highlighted in **Figure 4.8**, detailed in **Table 4.2**) generated unexpectedly low % genome coverage (49–73%) given the RT-qPCR Ct values (22–29) but were not part of sequencing runs 30 or 31. Sequences from all three samples were fragmented throughout ORF 1, with ORF 2 absent downstream from the capsid protruding domains and absent ORF 3 (**Figure 4.9**). In all three cases, the % OTR (0.01, 2.53 and 6.76%) and average read depth (1-, 120- and 137-fold) was low for ORF 1 despite apparently good Ct values. Coverage of ORF 1 and the 5' end of ORF 2 was sufficient to confirm two samples as GII.4 and one as GII.5 using the norovirus genotyping tool; we have shown good sequencing outcomes for both genotypes in other samples (**Table 4.1** on Page 121). It is not possible to exclude the possibility of a novel recombinant strain, with recombination at the P1/P2 junction in ORF 2, and

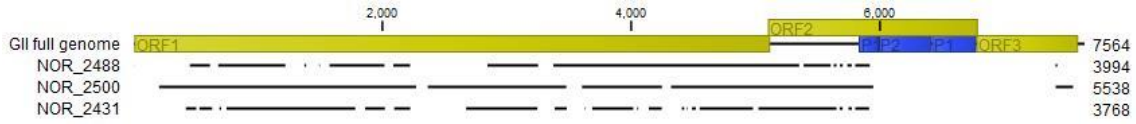
subsequent failure due to missing complementary baits in the enrichment; however if this were the case we would expect to see good coverage of the enriched region, in this case ORF 1, which we do not. Moreover all three samples had been re-extracted at referring centres and the Ct value supplied was obtained from PCRs carried out on the original diagnostic extracts. This, combined with the low coverage of ORF 1, suggests that extraction failure at the local hospital may explain the unexpected sequencing failure. It has not been possible to test either possibility, since none of the original sample remains.

Table 4.2 Sequencing results of samples with RT-qPCR Ct value <30 and unexpectedly low % genome coverage (<80%), excluding samples from sequencing runs 30 and 31 (n = 3; highlighted in Figure 4.8)

Sample name	Genotype	RT-qPCR Ct value	% genome coverage	% OTR*	Average read depth
NOR_2431	GII.5	22	49%	0.01%	1
NOR_2488	GII.4	29	52%	6.76%	137
NOR_2500	GII.4	22	73%	2.53%	120

*% OTR, % on-target-reads

Figure 4.9 Alignment of samples with RT-qPCR Ct value <30 and unexpectedly low % genome coverage (<80%), excluding samples from sequencing runs 30 and 31 (n = 3; highlighted in Figure 4.8 and detailed in Table 4.2)



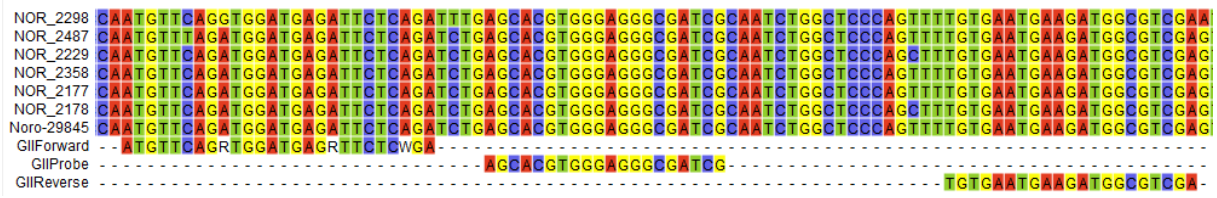
The outliers discussed above were assumed to have failed due to the aforementioned technical problems. Six of the samples from Runs 30 and 31 had sufficient residual specimen to be re-extracted and re-sequenced; all of these were successfully sequenced on repeat, which suggests the assumed reason for failure is correct. However, since there was no residual specimen to repeat the remaining outliers, the possibility that these were sequence failures due to other limitations with the sequencing pipeline, rather than the aforementioned technical problems, cannot be excluded.

One remaining outlier had only 13% genome coverage with a PCR Ct value of 38. This sample was not part of Runs 30 or 31. The possibility that this sample had poor genome coverage due to low virus titre (a Ct value of 38 is near the limit of detection for PCR) cannot be excluded, therefore this sample was not considered an outlier and was included in the calculation of predicted genome coverage.

4.3.4 Low titre samples

Seven samples generated full genome sequences in spite of low viral titres (PCR Ct ≥36). To determine whether these samples had misleadingly late Ct values due to a mismatch in the RT-qPCR primer target region, the seven genome sequences were aligned to the RT-qPCR primer and probe sequences used to generate the Ct value. There were no mismatches in the primer or probe sites (**Figure 4.10**), suggesting they are genuinely low titre samples and confirming the sensitivity of the method for low titre samples.

Figure 4.10 Nucleotide alignment of all samples with real-time PCR Ct value ≥36 (i.e. apparently low titre) but >80% genome coverage (n = 7) by SureSelect target enrichment sequencing



Degenerate nucleotide code: R, A/G; W, A/T

4.3.5 Comparison to capsid genotyping

96% (158/164) and 100% (164/164) of samples processed in parallel were successfully genotyped by PCR of the capsid shell domain with Sanger sequencing and by our method (SureSelect target enrichment), respectively (**Table 4.3**). For the 158 samples typed by both methods, there was 100% agreement in the respective genotypes. Of the 6 samples that failed capsid typing by PCR, four were GII.4, one GII.7 and one GI.3 (**Table 4.4**).

Table 4.3 Genotyping success of samples processed in parallel by SureSelect target enrichment for full genome sequencing and PCR amplification of capsid shell domain followed by Sanger sequencing

	Full genome sequenced	**Full genome failed****	**Total**
PCR amplified	158	0	158 (96%)
PCR failed*	6	0	6
Total	164 (100%)	0	164

* no amplification by PCR

** <90% genome coverage and <100-fold read depth

Table 4.4 Sequencing results of samples that failed to amplify the capsid shell domain by PCR for genotyping, but were successfully sequenced by SureSelect target enrichment (n=6)

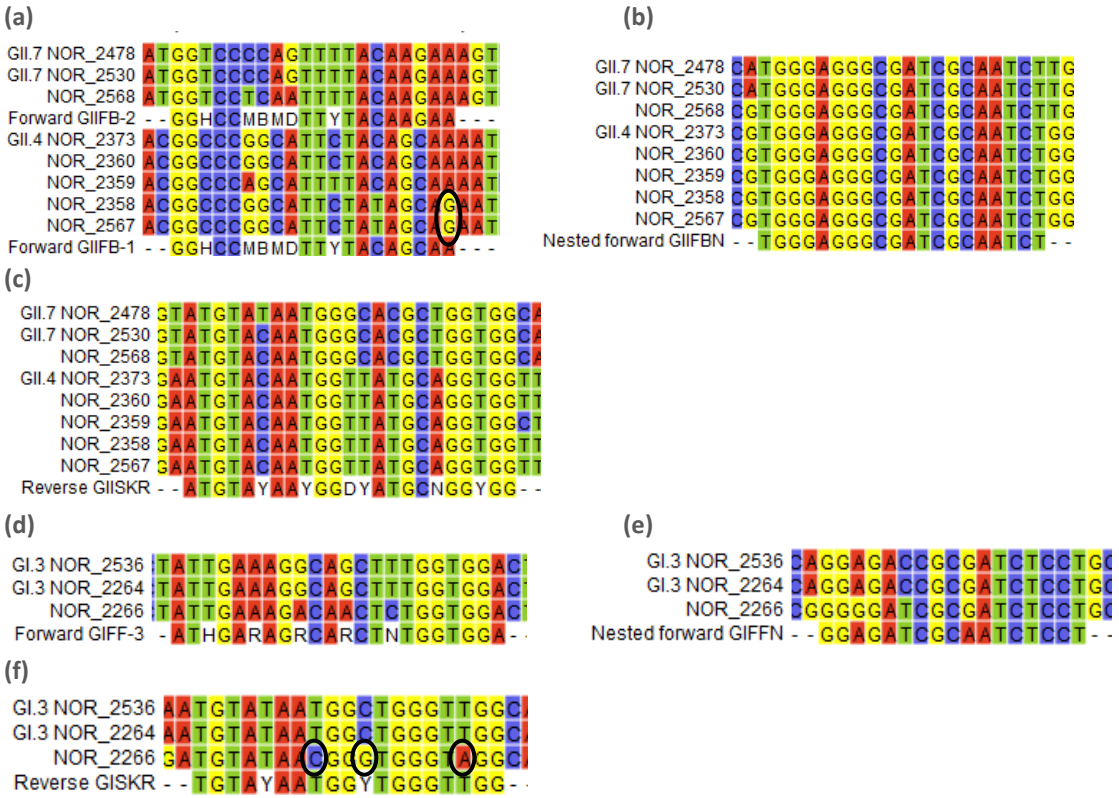
Sample name	Genotype	RT-qPCR Ct value	% genome coverage	% OTR*	Average read depth
NOR_2266	GI.3	20	100%	70%	13,957
NOR_2568	GII.7	31	99%	83%	11,235
NOR_2567	GII.4	27	100%	91%	23,068
NOR_2360	GII.4	33	100%	15%	3,995
NOR_2359	GII.4	33	100%	8%	2,281
NOR_2358	GII.4	37	99%	55%	11,740

*% OTR, % on-target reads

Two of the failed samples, with Ct values 20 and 27, had mismatches at the genotyping PCR primer sites (**Figure 4.11**) which accounts for genotyping failure in these instances.

The remaining four of the six samples that failed genotyping had Ct values >30 (range 31–37), which suggests the genotyping PCR is less sensitive than sequencing by target enrichment.

Figure 4.11 Nucleotide alignment of all GII (n=5) and GI (n =1) samples that failed to amplify the capsid shell domain by PCR for genotyping, but were successfully sequenced (>99% genome coverage) by SureSelect target enrichment. Additional sequences (labelled with genotype) that were genotyped successfully are included for comparison. Mismatches predicted to cause genotyping PCR amplification failure are circled. **(a)** GII forward; **(b)** GII nested forward; **(c)** GII reverse; **(d)** GI forward; **(e)** GI nested forward; **(f)** GI reverse



Degenerate nucleotide code: H, A/C/T; M, A/C; B, G/T/C; D, G/A/T; Y, C/T; R, A/G; N, G/C/A/T

4.3.6 Contamination

Two “negative extract” samples, consisting of Buffer ASL that was treated in the same way as, and alongside, stool samples, were negative for norovirus RNA by RT–qPCR. Nonetheless target enrichment and sequencing generated 16–36% OTR with 3–81-fold read depth (**Figure 4.12 a**). The genome coverage for each sample was only 9 and 12% (**Figure 4.12 b**), with reads fragmented across the genome (**Figure 4.13**). The mapped regions do not correspond to PCR amplicon sites.

Figure 4.12 Distribution of **(a)** mean read depth and **(b)** % genome coverage for stool samples and negative extracts. Red lines indicate median values

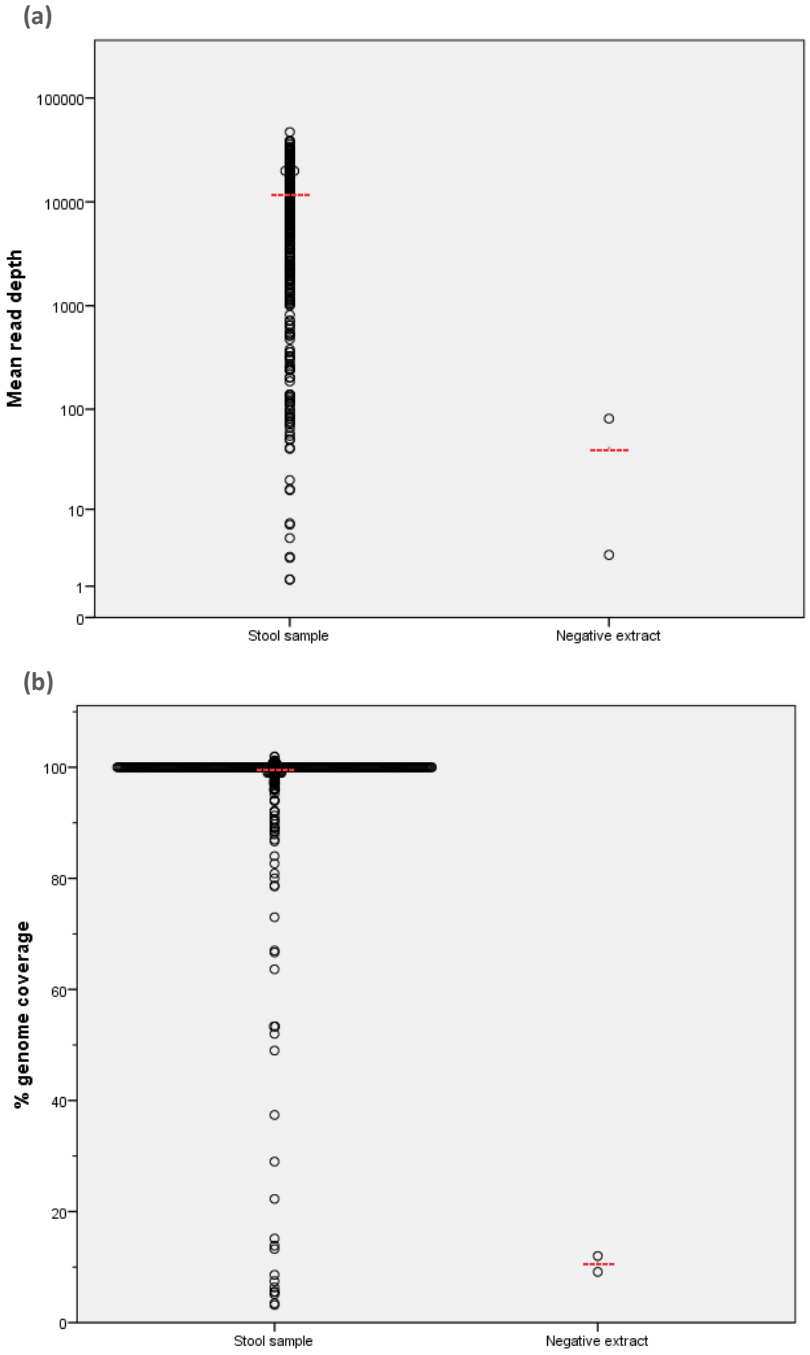


Figure 4.13 Contigs generated from Negative Extract mapping to norovirus full genome

4.3.7 Mixed infections

Three (3/507) samples were identified as having sequences from more than one genotype during the assembly pipeline (**Table 4.5**). For two of the samples, the mixed infections were evident during the "mapping to reference list" step of the *de novo* pipeline (**Figure 2.2** on Page 70 in General Methods), in which reads mapped to reference sequences corresponding to multiple norovirus genotypes, as per **Table 4.5**. For the third sample, mixed infection was evident during the "align contigs to single reference of appropriate genotype" step, in which a full length contig mapped to the reference sequence at ORF 1 but not at ORF 2 and ORF 3.

Table 4.5 Mixed infections in clinical specimens identified during assembly pipeline

Sample name	Genotypes present	Step in assembly pipeline that identified mixed infection	Number of reads mapping to each genotype in filtering step
NOR_2565	GII.3, GII.4 and GII.2	Filtering step (mapping to reference list)	GII.3: 792,264 GII.4: 113,068 GII.2: 859,899
NOR_2276	GII.3 and GII.6	Filtering step (mapping to reference list)	GII.3: 425,961 GII.6: 373,439
13V35152	GII.3 and GII.4	Align contigs to reference	GII.4: 470,377 GII.3: 1,155

Comparison of the consensus sequences generated from a single infection and from a simulated mixed infection showed 178–332 single nucleotide polymorphisms (SNPs) and 95.53–97.61% sequence identity between the consensus sequences from the single and mixed datasets (**Table 4.6**).

Table 4.6 Comparison of consensus sequences generated from de novo assembly in single infections (original) and a simulated mixed infection (mixed)

	Samples data originates from	Number of reads mapping	Read depth	Consensus sequence length	% identity between consensus sequences	SNP difference between consensus sequences	SNPs in ORF1	SNPs in ORF2	SNPs in ORF3
GII.3 original	Noro-14069	413,812	3,154	7,555	97.61	178	163 /178	15 /178	0 /178
GII.3 mixed	Noro-14069 & -28464	630,742	3,850	7,459					
GII.4 original	Noro-28464	431,263	2,917	7,489	95.53	332	284 /332	22 /332	26 /332
GII.4 mixed	Noro-14069 & -28464	676,523	4,017	7,426					

4.3.8 Turn-around times and costs

During a 5 week trial of sequencing in real-time at GOSH, actual turn-around times, from specimen receipt in the laboratory to sequencing result, were three weeks. This is based on weekly batching of specimens. The first week comprised batched RNA extraction of all norovirus positive specimens for that week, the second week cDNA synthesis and the third week SureSelect library preparation and sequencing. For library preparation reagents to remain cost effective samples must be sequenced in batches of 12 samples all with the same target pathogen; during the five weeks of prospectively sequencing in real-time this was only achieved in two of the five weeks, with one, five and eight samples in the other three weeks.

If one assumes no delays for batching, the best possible turn-around times associated with full genome sequencing by SureSelect target enrichment from RNA to sequencing result is 6 days; three days longer than genotyping (RNA-dependent RNA polymerase and capsid regions) by PCR and Sanger sequencing. SureSelect target enrichment incurs an extra cost of £54 when reagents are purchased in bulk (**Table 4.7**).

Table 4.7 Turn-around times and costs associated with norovirus genotyping by PCR and Sanger sequencing compared to SureSelect target enrichment full genome sequencing

Genotyping method	Hands on time	Total turn-around time	Reagent costs per sample
PCR and Sanger sequencing*	7 hours	3 days	£32
Full genome sequencing by SureSelect target enrichment	11 hours 30 minutes	6 days	£86–£93**

* PCR amplification of three sites of interest for norovirus genotyping; RNA-dependent RNA polymerase (RdRp), capsid shell domain and capsid P2 domain, including one round of nested PCR, assuming RdRp and capsid shell domain targets are amplified and sequenced simultaneously

** Cost based on batches of 96 or 48 samples and sequencing on an Illumina MiSeq.

4.4 DISCUSSION

Target enrichment is an effective method for sequencing norovirus full genomes across genotypes with a high read depth averaging over 12,000-fold and complete or almost complete genomes in 89% of samples. Target enrichment achieved median genome coverage of 100% across all sequenced samples and, once outliers were accounted for, over 80% genome coverage regardless of the viral titre.

4.4.1 Analysis of mixed infections

The *de novo* assembly pipeline developed as part of this study identified mixed genotype infections in three samples. However with as many as 332 SNPs between the consensus sequences generated from a single and simulated mixed infection, a reliable consensus sequence cannot be generated using this assembly pipeline. This is due to miss-mapping of reads in relatively conserved regions, as evidenced by the majority of SNPs being found in ORF1 (163/178 and 284/332 in the GII.3 and GII.4 consensus sequences respectively). Thus whilst this pipeline can identify infections with a mixture of genotypes, an alternative approach is required for assembly and generating the consensus sequence, possibly involving the use of minority variants and haplotype reconstruction.

It is possible that samples identified as mixed infections during the assembly pipeline, due to reads mapping to multiple genotypes, are sequencing artefact rather than true mixed infections. Given that contaminating norovirus reads were detected in the two "negative extracts", it is possible that contaminating reads from other samples mimic a mixed infection, when there is not one. This may be the case for sample 13V35152 (**Table 4.5**) in which only 1,155 reads mapped to GII.3 compared to 470,377 reads mapping to GII.4; however the GII.3 reads generated a 3,000 nucleotide contig mapping to ORF 2, ORF 3 and the 5' end of ORF 1 whereas the contaminating reads in the negative extracts were fragmented across the genome (**Figure 4.13**). Nonetheless in this instance further evidence is required to conclude there is a true mixed infection, such as re-extracting and re-sequencing the sample to replicate the results. This was not possible as there was no residual specimen for repeat processing.

4.4.2 Comparison to other methods for whole genome sequencing

Previous reports have described whole norovirus genome sequencing with overlapping PCR amplicons or using RNASeq, the findings of which are summarised in **Table 4.9**. PCR- based methods yield high read depth; however, due to sequence heterogeneity

between genotypes, primers generally need to be genotype specific [148]. Broad-range primers have been reported by Cotton *et al.* [149] nonetheless this approach retains a limited success rate; full genome sequences were amplified from a comparable proportion of samples of GII.13, GII.6 and GII.4. However PCR fared worse, recovering fewer full genomes from GII.2, GII.3, GII.7 and GI (**Table 4.8**). It may be possible to design broad-range primers that have greater sequencing success across genotypes; however there are currently no alternatives published in the literature. Norovirus whole genome sequencing from a single 7.5 kb amplicon has also been described and used to generate 25 full genome sequences [146] however the authors do not report the success rate using this approach; it is generally very difficult to amplify fragments of such a size. Conversely this chapter reports complete or nearly-complete genome sequences in 93% of processed samples. In target enrichment, baits are designed using all publically available norovirus sequences, across all GI and GII genotypes; unlike PCR which uses a single primer at each target site, multiple baits are designed to cover each position in the genome thus accounting for sequence variation between norovirus genomes. This potentially allows un-biased sequencing across known genotypes in a single reaction. A disadvantage of the method is that it may fail to generate sequences for a newly emerging genotype where the existing baits are a poor match.

Table 4.8 Percentage of samples belonging to each genotype for which full genomes were achieved in this study, using target enrichment, and by Cotton *et al.* [149] using PCR

Genotype	This study	Cotton *et al.* (2015)
GII.13	100%	83%
GII.6	95%	88%
GII.4	89% (93% irrespective of read depth)	92%
GII.2	88%	40%
GII.3	87% (90% irrespective of read depth)	77%
GII.7	90%	0%
GI	100%	20%

Whole transcriptome sequencing, or direct sequencing of total RNA (RNASeq), involves sequencing the total RNA or mRNA content of a stool specimen. The advantage of RNASeq is that there is no requirement for PCR primers therefore it is completely unbiased. Although whole genomes by RNASeq reported to date are predominantly GII.4, it is theoretically possible to sequence all genotypes with equal success as evidenced by Bavelaar *et al* who successfully sequenced five non-GII.4

genomes [189]. The data generated by RNASeq is sufficient to generate almost complete norovirus genome sequences; 40–100% of reported samples achieved >90% genome coverage [147, 178, 179, 189] (summarised in **Table 4.9**). However the median % OTR across all reported samples is only 2–3% using a MiSeq or HiSeq [178, 179] and 28% using an Ion Torrent PGM [189], compared to 81% OTR by SureSelect target enrichment. The proportion of on-target reads using target enrichment is not 100% meaning that, as with RNASeq, non-target sequences are generated; however the % OTR of target enrichment is, with the exception of PCR-based methods, superior to other methods that are currently published in the literature. The high proportion of non-target data using RNASeq makes the technique uneconomical and, critically, results in low read depth; on average only 9–259-fold using a MiSeq or HiSeq [147, 178, 179] and 1,309 using an Ion Torrent PGM [189]. Conversely, the median read depth using target enrichment is over 12,000-fold which allows large batches of samples to be sequenced on a single MiSeq run and downstream analysis of minority variants.

Table 4.9 Summary of norovirus whole genome sequencing reports in the literature

Author	Country	Method	Sequencing platform	Assembly	Number of samples	Genotypes (% success*)	% OTR	Genome coverage	Read depth
Chhabra (2010) [190]	India	Overlapping PCR (9 amplicons)	Sanger sequencing	Not stated	3	GII.4 and GII.b_GII.3†	n/a**	100%	n/a††
Won (2013) [191]	South Korea	Overlapping PCR (10 amplicons)	Sanger sequencing	Not stated	1	GII.P12_GII.13†	n/a**	100%	n/a††
Eden (2013) [146]	Australia	Long PCR (7.6kb)	Sanger sequencing	Not applicable	25	GII.4†	n/a**	100%	n/a††
Kundu (2013) [148]	UK	Overlapping PCR (22 amplicons)	Roche 454	De novo	13	GII.4 †	n/a**	86–99%	Median 580 (423–951)
Cotten (2014) [149]	Vietnam	Overlapping PCR (3 amplicons)	Illumina MiSeq	De novo	265	GI (20%) GII.2 (40%) GII.3 (77%) GII.4 (92%) GII.6 (88%) GII.7 (0%) GII.9 (100%) GII.12 (50%) GII.13 (83%) Overall success 78%	n/a**	100%	Not stated
Botha *et al.* (2016) [192]	South Africa	Overlapping PCR (3–5 amplicons)	Sanger sequencing	Not stated	11	GII.4	n/a**	100%	n/a††

Author	Country	Method	Sequencing platform	Assembly	Number of samples	Genotypes (% success*)	% OTR	Genome coverage	Read depth
Nakamura (2012) [179]	Japan	Whole transcriptome amplification kit (70 cycles PCR)	Roche 454	Reference mapping	5	GII.4 (40%)	Median 3% (0.05–60%)	Median 84.5% (2.1–98%)	9–259
Batty (2013) [178]	UK	mRNA RNASeq	Illumina MiSeq and HiSeq	Reference mapping	3 (MiSeq) plus 77 (HiSeq)	GII.4 (99%)	MiSeq: median 1.8% (0.12–1.90) HiSeq: median 2.7% (0.01–97.98%)	MiSeq: 97–99% HiSeq: Mean 97% (59 – 99%)	MiSeq: Not stated HiSeq: Median 100 (10–1,000)
Wong (2013) [147]	UK	mRNA RNASeq	Illumina MiSeq	De novo	32	GII.4 (66%)	Not stated	21/32 >97%	Not stated
Bavelaar *et al.* (2015) [189]	Netherlands	Whole transcriptome amplification with ribosomal RNA depletion	Ion Torrent PGM (318 chip)	Reference mapping	10	GI.1 (100%) GI.6 (100%) GII.4 (100%) GII.6 (100%) GII.21 (100%) GII.2 (100%)	Median 28% (0.7–70.9%)	100%	Median 1309 (25–3607)

*>90% genome coverage

** Not applicable for PCR amplicon sequencing

† Success rate not reported

†† Read depth not applicable for capillary sequencing

4.4.3 Comparison to PCR-based genotyping

Target enrichment is shown in this chapter be superior to PCR capsid amplification for genotyping; all samples (164/164) that were processed in parallel successfully generated genome sequences by target enrichment, whereas 96% (158/164) were successfully amplified by capsid typing PCR. Four of the six samples that failed capsid genotyping but were sequenced by target enrichment had low norovirus titres (based on PCR Ct values), which suggests target enrichment is more sensitive than the conventional genotyping methods. The remaining two failed samples had primer mismatches that account for amplification failure. Target enrichment overcomes the limitations of primer design by allowing multiple baits with different sequences to target each region of the genome, thus accounting for sequence heterogeneity in a way that PCR primers cannot.

Unlike classical genotyping, which requires sequential PCR and sequencing reactions yielding only fragments of the genome in return, full genome sequences can, in a single reaction, provide us with the RNA polymerase and capsid sequences, which are important for genotyping, and in addition can identify recombination and reveal minority variants in the intra-host viral population.

4.4.4 Quality assurance

Despite good molecular practice, low level contamination does occur in the sequencing pipeline described in this chapter. Since negative extracts were RT-qPCR negative but target enrichment yielded reads that map to the norovirus genome, the suspected source of contamination is the automated equipment used for target enrichment and sequencing library preparation. In the context of norovirus-positive specimens, the contamination is low; reads are fragmented and only map to 9-12% of the genome with <100-fold read depth, which is significantly below the observed median % genome coverage and read-depth seen for norovirus-positive samples (100% and >12,000-fold, respectively) and below the 95% prediction intervals for % genome coverage (92–100% for a sample with Ct <40). These findings support the proposed acceptance criteria for downstream analysis, which is >100-fold read depth and >90% genome coverage. Due to the potential for low level contamination, specimens for which norovirus RNA is not detectable by real-time PCR should not be sequenced.

For robust quality assurance, which is required for diagnostic laboratory accreditation, a negative and positive control should typically be included in every run. The negative control should be a norovirus-negative specimen or buffer that has been co-extracted

and processed with the norovirus-positive specimens and the positive control a known norovirus-positive specimen. As described above, only norovirus positive samples should be processed for sequencing; these act as their own positive control therefore a separate positive control on every run is not strictly required. However a positive control with each batch would be necessary to ensure enrichment has performed as expected (for example expected read depth given the batch size). The positive control is of particular importance if one plans to identify minority variants; in this instance the control should include minority variants of known frequency. Implementation of acceptability criteria, as described here, overcomes problems of low-level contamination in the system, however this does not account for gradual accumulation of contamination in the system or gross contamination of reagents. For this reason a negative control should be included in sample processing. Ideally this should be with every batch; however in the context of low-throughput processing in a diagnostic laboratory the cost of this may be prohibitive.

4.4.5 Cost of sequencing

The cost of target enrichment whole genome sequencing is around £50 more expensive than PCR genotyping of the capsid and polymerase genes. Whole genome sequencing using overlapping amplicons is comparable in cost to target enrichment; therefore target enrichment is not more expensive than other whole genome sequencing methods. However there is an economy of scale with sequencing, including the library preparation reagents but in particular the sequencing run; an Illumina MiSeq sequencing cartridge costs £685 regardless of whether one or 100 samples are sequenced. Therefore the costs discussed here were only achieved by sequencing in batches of 48 or 96 samples per run. This is achievable for a large study, or for a regional sequencing service which can fill a run with other samples, however is not realistic for an individual diagnostic laboratory.

4.4.6 Turn-around times

Once RNA is extracted, turnaround time for target enrichment is 6 days compared to three days for capsid and polymerase genotyping by PCR, including sequencing time. The semi-automated target enrichment hands-on-time is 4 hours more than conventional genotyping and comparable to RNASeq. It should be possible to synthesise cDNA, prepare the sequencing library and complete a sequencing run within 6 days, however in reality this took two weeks as cDNA synthesis was batched one week and library prep and sequencing batched the following week. In addition there is a one week lag between specimen collection and cDNA synthesis; this is due

to waiting for the diagnostic real-time PCR result to identify norovirus positive samples followed by weekly batching of RNA extraction for sequencing. Therefore, over all, there was a three week turn-around time from specimen collection to result. This three week turn-around time was only made cost-effective by topping up weekly batches to 12 specimens with archived retrospective samples, and by filling MiSeq runs to 48–96 samples using samples with other target pathogens from other studies. If other samples were not available, sequencing in real-time would either not have been cost effective, or turn-around times would have been extended by several weeks to enable a full run.

The utility of norovirus full genome sequencing in real-time is currently limited by drawbacks of batch processing. These would be in part addressed by random-access processing and shortened library preparation hybridisation times, which are the subject of future work. In addition, to optimise turn-around times and cost-effectiveness, norovirus whole genome sequencing needs to occur in a laboratory with a high-throughput of next generation sequencing. In time this may be individual diagnostic laboratories, but at the moment this would have to be a regional laboratory or sequencing service.

4.4.7 Conclusions

Target enrichment is a robust and sensitive method for sequencing norovirus full genomes that overcomes the limitations of other published methods, namely of primer design in PCR and of non-target sequencing in RNASeq. It is comparable in cost to other whole genome sequencing methods, and not significantly more expensive then PCR-based genotyping despite providing far greater amounts of data; however batch sequencing is necessary to achieve cost-savings, which has implications on turn-around times and thus clinical utility. The read depth and breadth of genotypes that can be sequenced is superior to other published sequencing methods.

Low level contamination occurs in the sequencing pipeline, however this is distinguishable from genuine data which has significantly greater genome coverage and read depth. Nevertheless only specimens for which norovirus is detectable by RT-qPCR should be sequenced.

4.5 ACKNOWLEDGEMENTS

Stool specimens for sequencing were from Great Ormond Street Hospital Virology department (n=326), Royal Free Hospital Virology department (n=24), Norfolk and

Norwich University Hospital Microbiology department (n=71) and Public Health England (PHE) Virus Reference Department (n=86).

Stool samples from Norfolk and Norwich and PHE were extracted by the respective referring laboratory. I extracted 144 stool samples from GOSH. The remaining 182 stool samples from GOSH and all samples from the Royal Free were extracted by Divya Shah, under my supervision.

cDNA synthesis for 144 samples was done by me. cDNA synthesis for the remaining samples was done by the UCL Pathogen Genomics Unit (PGU).

Bait design, sequencing library preparation and Illumina sequencing was done by the UCL PGU.

Full genome assembly of 201 samples and one Negative Extract was done by me. Assembly of the remaining 306 samples and second Negative Extract was done by the UCL PGU.

Statistical advice was provided by Eleni Pissaridou, from the Institute of Child Health (ICH) Statistical Support Services for the calculation of predicted % genome coverage, including transformation of % genome coverage for the linear regression model and exclusion of outliers attributed to pre-sequencing technical failures.

CHAPTER 5

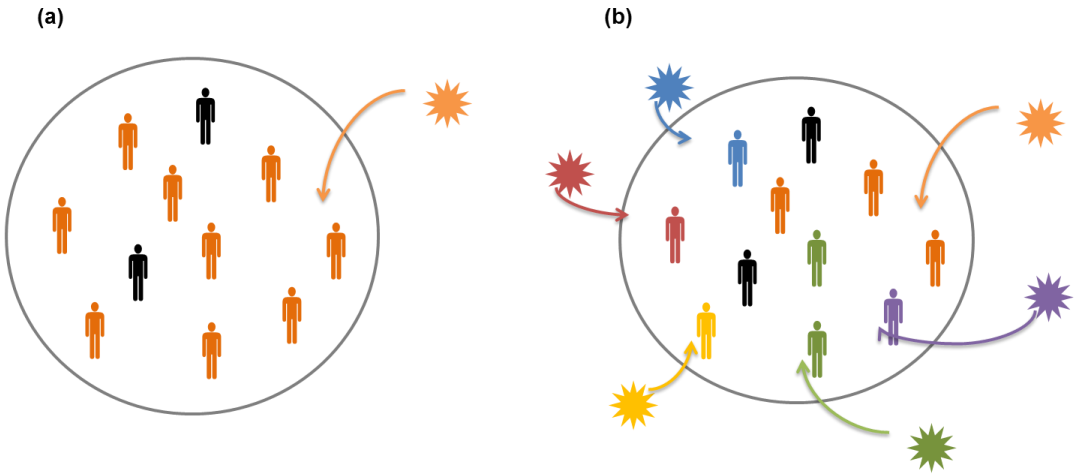
NOROVIRUS MOLECULAR EPIDEMIOLOGY IN A PAEDIATRIC HOSPITAL: UTILITY OF FULL GENOME SEQUENCING IN A CLINICAL SETTING

5.1 INTRODUCTION

Molecular epidemiology allows us to compare either complete or partial viral genome sequences between infected cases; based on phylogenetic analysis one can infer shared ancestry between infecting viruses. This inference can allude to transmission events between linked infections or, conversely, exclude cases from an outbreak.

Given that the number of cases of norovirus characteristically peaks during winter, it can be difficult to distinguish a true outbreak occurring in an institution, with introduction of a single strain and consequent spread, from multiple introductions of unlinked cases that happen to occur contemporaneously (**Figure 5.1**). It is important to distinguish between these two scenarios for purposes of infection control; the first scenario suggests a break-down in infection control practices and requires interventions to interrupt transmission, whist the second scenario suggests high rates of norovirus circulation but with effective infection control practice limiting transmission.

Figure 5.1 Two scenarios of norovirus nosocomial epidemiology; **(a)** single introduction of one strain which spreads throughout the hospital thus causing the majority of detected cases, or **(b)** multiple introductions of several different strains with limited transmission of each strain



Norovirus outbreaks in healthcare settings are characteristically dominated by a single genotype, typically GII.4 [128]. Genotyping of norovirus infections allows patients with a different genotype to be excluded from the outbreak; however it cannot confirm or exclude transmission between patients infected with the same genotype. In this instance molecular epidemiology using phylogenetic analysis may be informative; in influenza transmission studies phylogenetic analysis demonstrated that epidemiologically related cases, i.e. linked in time and place, did not always share

related viruses [193]. The lack of diversity in the polymerase and capsid shell domains used for norovirus genotyping means these regions provide insufficient information for tracking outbreaks [194]. Consequently these regions cannot be used to investigate transmission between infections caused by the same genotype.

To date, reports of norovirus molecular epidemiology have largely utilised partial genome sequences, as these can be amplified by PCR and Sanger sequenced. In lieu of full genomes, due to its high level of diversity the P2 region of the capsid is the most informative region in norovirus outbreaks [145]. In immunocompetent people the P2 region is expected to be 100% identical between patients in whom a transmission event has occurred [140, 143], with a 10% probability of 1–2 nucleotide changes in samples collected three weeks post-infection [142]. Therefore in acute norovirus outbreaks amongst immunocompetent persons P2 sequencing has been used to identify independent clusters of transmission among patients infected with the same genotype and identify transmission events which were missed based on traditional epidemiological data alone [139, 140, 142-144]. Reports of norovirus P2 domain sequencing used for nosocomial molecular epidemiology are summarised in **Table 5.1**.

In immunocompromised patients, however, transmission may occur after *in vivo* evolution therefore single nucleotide polymorphisms (SNPs) are observed in the capsid sequence between linked patients [144]. Holzknecht *et al*. used PCR and Sanger sequencing to sequence the norovirus capsid sequence (1412 nt, including P2 domain) from 13 immunocompromised patients from an outbreak on a haematology ward in 2008 in Denmark and found up to 13 SNPs between linked patients [144]. Diversity in the capsid region in immunocompromised patients has the potential to make interpretation of phylogenies based on P2 sequences alone challenging as the previously employed rule of identical P2 sequences between linked cases does not apply.

To date, there are no comprehensive reports in the literature of norovirus molecular epidemiology in an immunocompromised population, neither using P2 sequences nor full genomes. This chapter aims to investigate the utility of norovirus whole genome sequencing in a nosocomial setting and to elucidate the transmission dynamics in a paediatric hospital with a large immunocompromised population.

Table 5.1 Summary of published reports of nosocomial norovirus molecular epidemiology using capsid P2 domain sequences

Authors	Method	Genotype	Year and country	Region analysed	Number of patients	Setting	Results
Xerry *et al.* 2008 **[143]**	PCR and Sanger sequencing	GII.4	2006–2007, UK	P2 (800 nt)	44	14 outbreaks in 2 hospitals	100% sequence identity within outbreaks Recommend ≥1 nt difference between patients suggests they are not linked
Xerry *et al.* 2010 **[195]**	PCR and Sanger sequencing	GII.3	2007, UK	P2	2	Paediatric primary immunodeficiency unit; stool from two chronically infected patients on the same ward plus 116 environmental swabs	Longitudinal patient sequences 98.6–99.8% sequence identity (3 “strains”) 3/18 environmental sequences identical to patient 15/18 environmental sequences not identical to patients; conclude unrelated to patients (dismiss intra-host evolution)
Sukhrie *et al.* 2011 **[140]**	PCR and Sanger sequencing	GII.3 and GII.4	2002–2007, Netherlands	P2 (600 nt)	105	Nosocomial infections in one hospital	100% sequence identity within clusters Clusters involved patients across several wards, previously not linked due to physical separation 5 clusters involved outpatients
Sukhrie *et al.* 2013 **[142]**	PCR and Sanger sequencing	GII.4	2009–2011, Netherlands	P2 (600 nt)	81	1 hospital outbreak and 3 nursing home outbreaks	Identical P2 sequences (0 SNPs) within clusters in 70/81 patients; 6/81 patients 1 SNP, 1/81 patients 2 SNPs
Holzknecht *et al.* 2015 **[144]**	PCR and Sanger sequencing	GII.4	2007–2008, Denmark	Capsid (1412 nt including P2)	55	Nosocomial infections in one hospital	7 distinct outbreaks identified, including interdepartmental transmission. 3/7 clusters: 0 SNPs between patients; 1/7 clusters: 1 SNP; 1/7 clusters: 2 SNPs; 1/7 clusters: 4 SNPs (suspect multiple introductions from community) 1/7 clusters (immunocompromised patients): 13 SNPs

5.2 MATERIALS AND METHODS

5.2.1 Stool samples for molecular epidemiology

The first positive stool sample of all new episodes of norovirus at GOSH Between 1st July 2014 and 17th February 2016 (19 months) were identified for inclusion in the study (n = 205). Ten patients had no residual stool sample and six were PCR negative on re-extraction. The first norovirus-positive stool samples from the remaining 189 patients were included in this study.

In addition, 71 samples were collected from Norfolk and Norwich University Hospital (NNUH) between January and April 2015. This represents a suspected norovirus outbreak over a two week period in March and April with 3 additional norovirus episodes from January and 3 from February 2015. Norfolk and Norwich University Hospital is a 1,200 bed general hospital mainly serving patients in Norfolk and North Suffolk. Whilst it provides paediatric services, it is not a paediatric hospital.

5.2.2 Longitudinal stool samples

94 longitudinal stool samples were sequenced from 25 patients, with two to seven samples per patient. Longitudinal samples were collected between November 2012 and January 2016.

5.2.3 Norovirus whole genome sequencing

Whole norovirus genomes were sequenced from all study samples by SureSelect target enrichment, as described in Chapter 2.

5.2.4 Sequence assembly

Reads were assembled using a *de novo* assembly pipeline, as described in Chapter 2, to generate a consensus sequence to be used in downstream analysis.

5.2.5 Phylogenetic analysis for molecular epidemiology

A full genome alignment and maximum likelihood tree was reconstructed for each genotype. Maximum likelihood phylogeny was also reconstructed using the P2 domain sequence of norovirus GII.4 episodes.

To assess the occurrence of recombination between genotypes, phylogenies were reconstructed for ORF1 sequences from samples with the same polymerase genotype but a different capsid genotype.

Details of all reconstructed phylogenies are described in Chapter 2.

5.2.6 Longitudinal analysis

A single maximum likelihood phylogeny was reconstructed for all longitudinal samples, as detailed in Chapter 2.

5.2.7 Genotyping

All full genome consensus sequences were submitted to the Norovirus Genotyping Tool (http://www.rivm.nl/mpf/norovirus/typingtool) [141] to determine the polymerase and capsid genotype.

5.3 RESULTS

5.3.1 Consensus sequences

In total, 184/189, 62/71 and 86/94 samples that were sequenced for molecular epidemiology at GOSH, NNUH and longitudinal analysis, respectively, generated greater than 90% genome coverage and >100-fold read depth, therefore were taken forward for downstream analysis. Twenty-two samples that generated less than 90% genome coverage and/or <100-fold read depth were excluded from analysis. One sequence was excluded from the molecular epidemiology phylogenetic analysis and two from longitudinal phylogenetic analyses as infection with a mixture of genotypes was detected during sequence assembly; the assembly pipeline used in this study cannot generate reliable consensus sequences from samples with multiple genotypes (as demonstrated in Chapter 4).

5.3.2 Patient immune statuses

Of the 184 patients included in molecular epidemiology analysis from GOSH, 110/184 (60%) are considered immunosuppressed based on the clinical specialty to which they were admitted (haematology, oncology, immunology, bone marrow transplant, solid organ transplant). The remaining 74/184 (40%) were medical or surgical patients, although some of these patients may have some degree of immunosuppression.

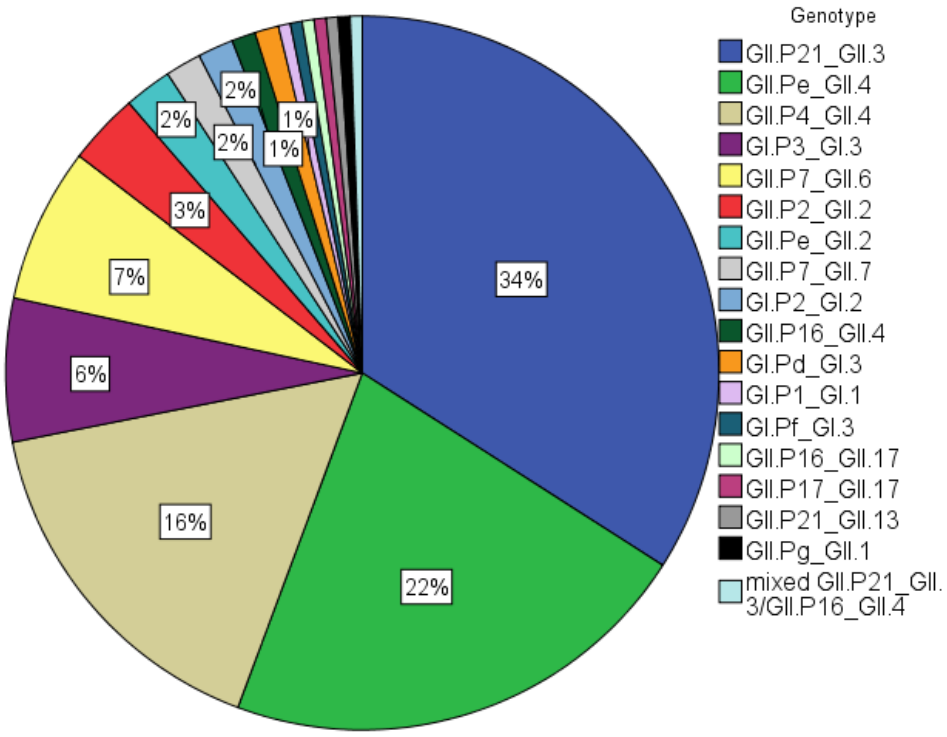
5.3.3 Genotypes

5.3.3.1 Genotypes using full genome sequences

The 19 month sampling period in this study built on the 12 month sampling period described in Chapter 3; as expected the 19 month sampling in this chapter

demonstrates a broad range of genotypes, with a third of sequenced samples (62/184, 34%) identified as norovirus capsid genotype GII.3 (**Figure 5.2**). Full genome sequences allow the polymerase and capsid genotype to be determined, identifying all of the GII.3 sequences as GII.P21_GII.3 (polymerase genotype II.21, capsid genotype II.3).

Figure 5.2 Norovirus genotypes identified in paediatric tertiary referral hospital (GOSH), July 2014 – February 2016 (n = 184)



Of the 73 samples with a GII.4 Sydney_2012 capsid only 40/73 (55%) had a GII.Pe polymerase as would typically be expected for a Sydney_2012 strain (GII.Pe_GII.4); 30/73 (41%) had a GII.P4 New Orleans_2009 ORF1 and 3/73 (4%) a GII.P16 ORF1 (GII.P4_GII.4 and GII.P16_GII.4, respectively). Multiple ORF1 genotypes were also identified with capsid sequences GI.3 (12/15 GI.P3_GI.3, 2/15 GI.Pd_GI.3, 1/15 GI.Pf_GI.3), GII.2 (6/10 GII.P2_GII.2, 4/10 GII.Pe_GII.2) and GII.17 (1/2 GII.P17_GII.17, 1/2 GII.P16_GII.17).

5.3.3.2 Recombination between genotypes

For each of the four ORF 1 maximum likelihood trees all of the ORF 1 sequences clustered according to their corresponding capsid genotype (**Figure 5.3**, **Figure 5.4**, **Figure 5.5** and **Figure 5.6**). This suggests there has been no recombination between genotypes.

Figure 5.3 Maximum likelihood phylogeny of ORF1 sequences from norovirus episodes with a **GII.P7** ORF1 genotype

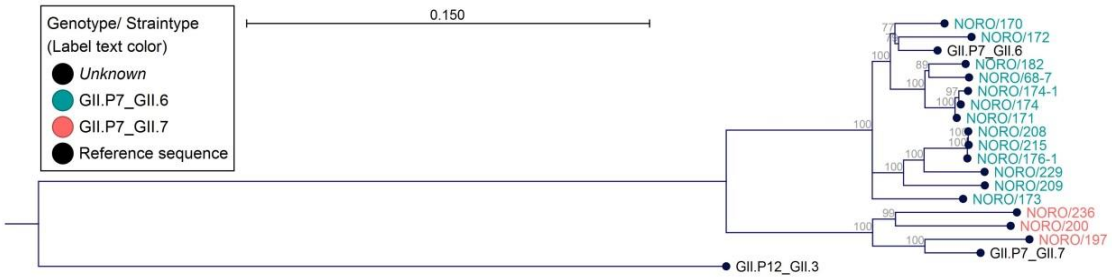


Figure 5.4 Maximum likelihood phylogeny of ORF1 sequences from norovirus episodes with a **GII.P16** ORF1 genotype

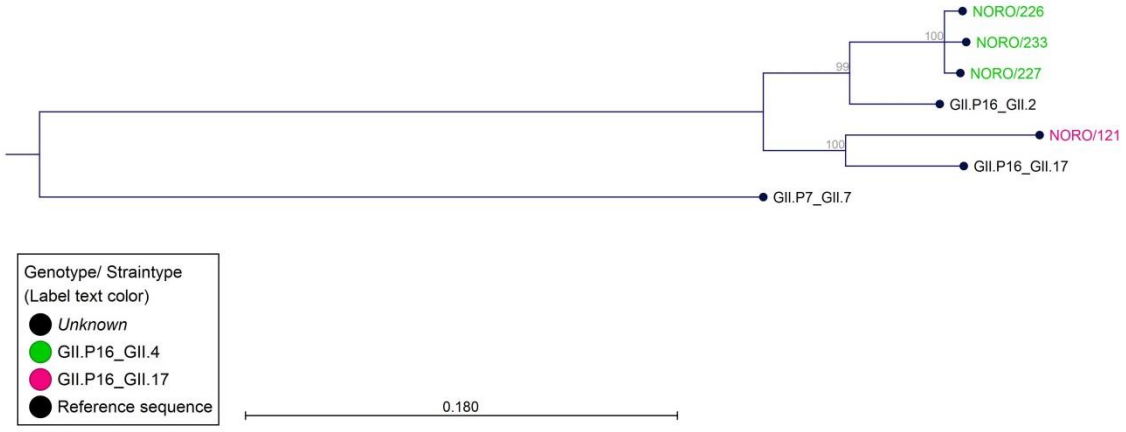


Figure 5.5 Maximum likelihood phylogeny of ORF1 sequences from norovirus episodes with a **GII.Pe** ORF1 genotype

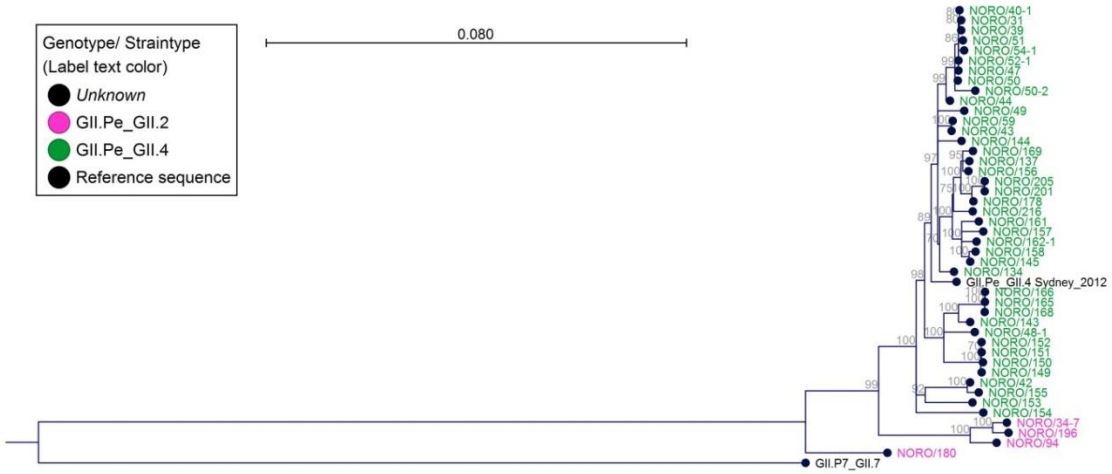
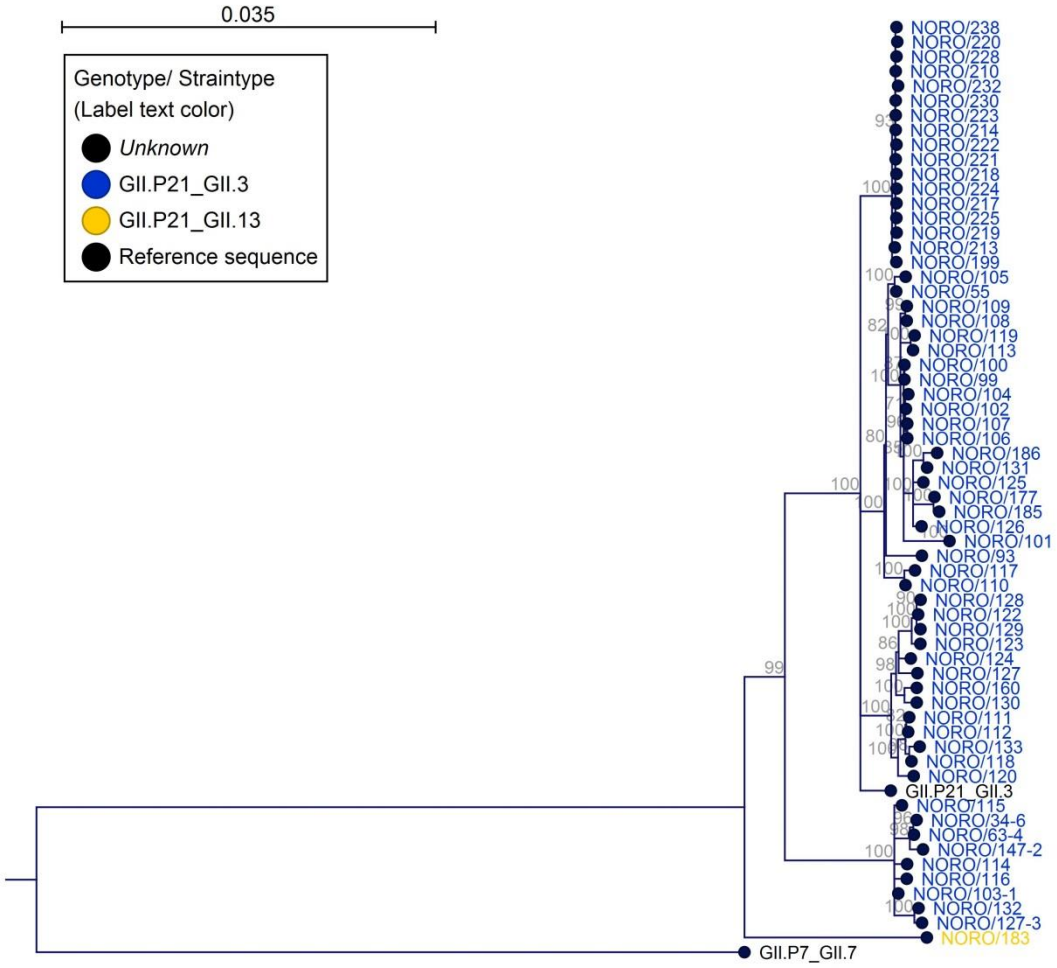


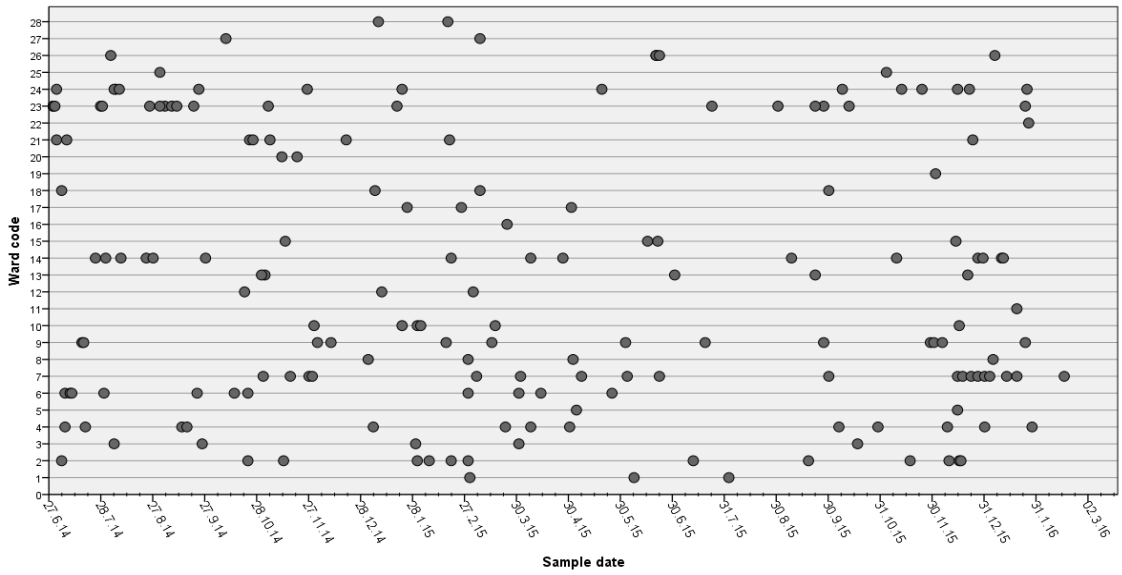
Figure 5.6 Maximum likelihood phylogeny of ORF1 sequences from norovirus episodes with a **GII.P21** ORF1 genotype



5.3.3.3 Genotyping for molecular epidemiology

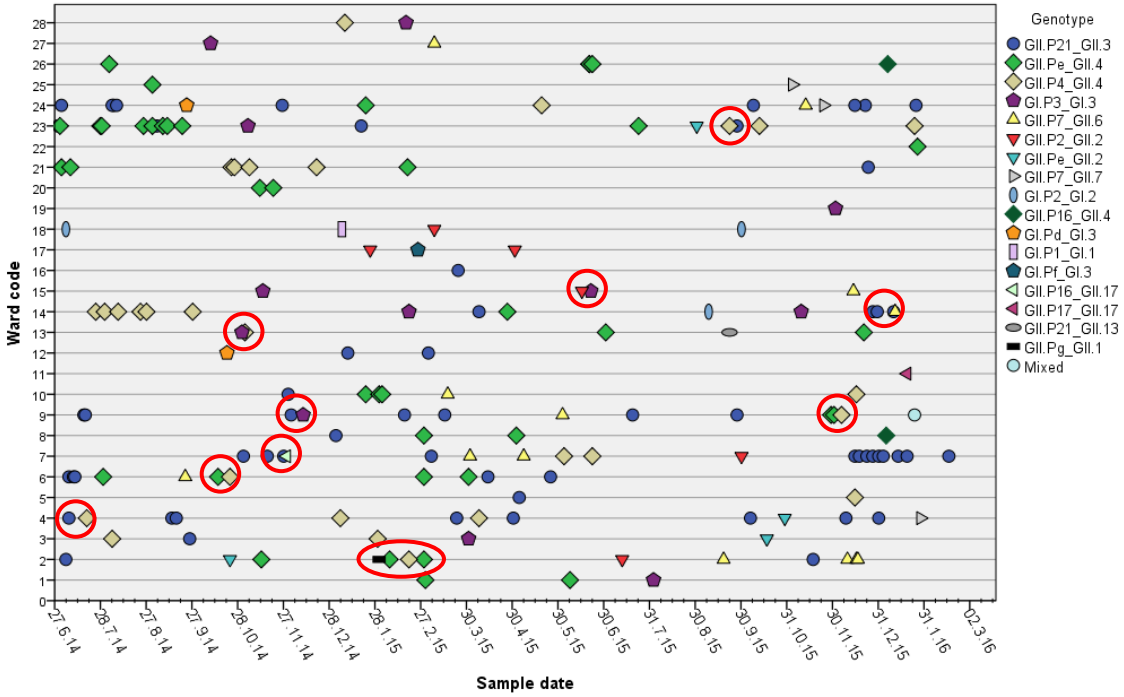
Norovirus infections at GOSH occur throughout the year and throughout the hospital (**Figure 5.7**). However it is not known whether infections are occurring due to transmission between patients or from independent sources.

Figure 5.7 Distribution of norovirus genotypes across wards in paediatric tertiary referral hospital (GOSH), July 2014 – February 2016. Each number on the y axis represents a different ward. Each point represents a new norovirus infection



Genotyping shows that several norovirus genotypes are co-circulating throughout the hospital at any given time (**Figure 5.8**). In ten instances patients were closely linked by classical epidemiology, i.e. time and place, but were infected with different genotypes (**Figure 5.8**, circled episodes) thus genotyping can exclude cross-transmission between these patients.

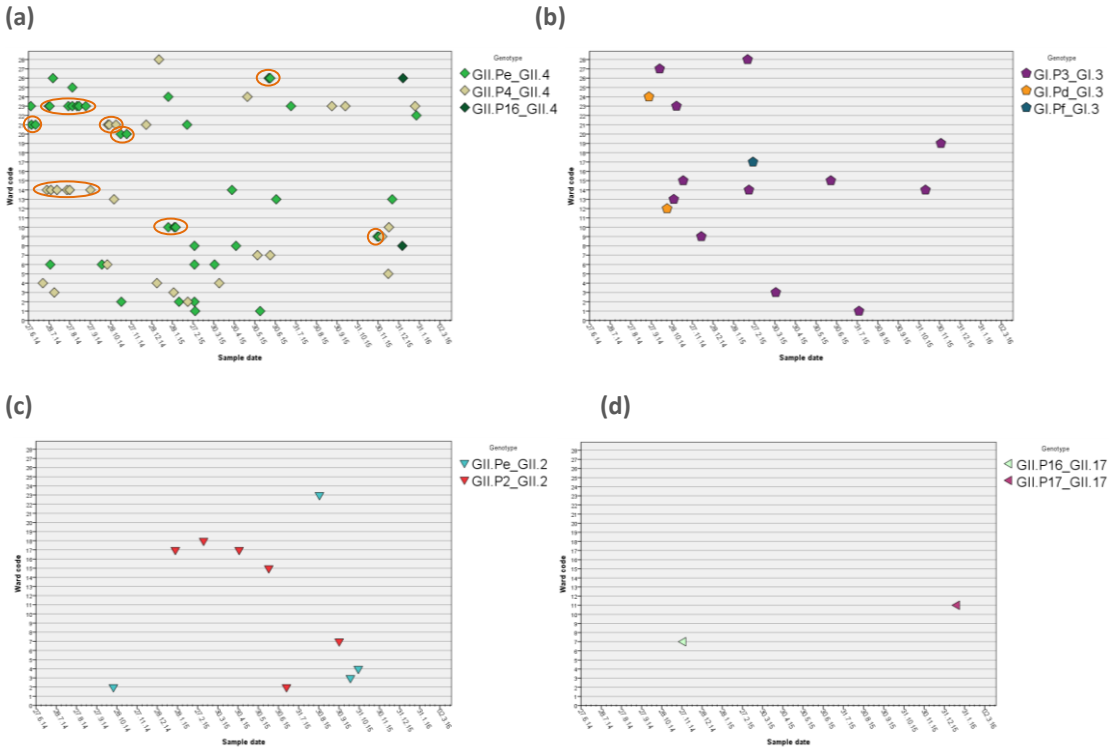
Figure 5.8 Distribution of norovirus genotypes across wards in paediatric tertiary referral hospital (GOSH), July 2014 – February 2016. Each number on the y axis represents a different ward. Each point represents a new norovirus infection, colour-coded according to infecting genotype. Red circles indicate norovirus episodes closely linked in time and place, but shown to be different genotypes



Genotyping using full genome sequences, as opposed to capsid sequences alone, increases the capacity to exclude cross-transmission in cases that share a capsid genotype but have a different polymerase genotype, as illustrated in **Figure 5.8** and **Figure 5.9**. In these instances episodes would have been indistinguishable by capsid genotyping alone.

Despite the utility of genotyping to exclude transmission in some instances, it is not possible with genotyping alone to ascertain whether norovirus episodes caused by the same genotype are linked. This is illustrated in **Figure 5.9 a**; the eight circled episodes highlight instances where transmission was possible as the episodes occurred on the same ward with identical genotypes. Similarly, there were eight instances where episodes caused by GII.P21_GII.3 occurred on the same ward at the same time, but cannot be differentiated by genotyping alone.

Figure 5.9 Distribution of norovirus genotypes across wards, demonstrating utility of polymerase genotype to discern norovirus episodes caused by the same capsid genotype. Each number on the y axis represents a different ward. Each point represents a new norovirus infection, labelled according to infecting genotype. **(a)** GII.4 capsid episodes (orange circles highlight episodes occurring on the same ward with the same genotype); **(b)** GI.3 capsid episodes; **(c)** GII.2 capsid episodes; **(d)** GII.17 capsid episodes

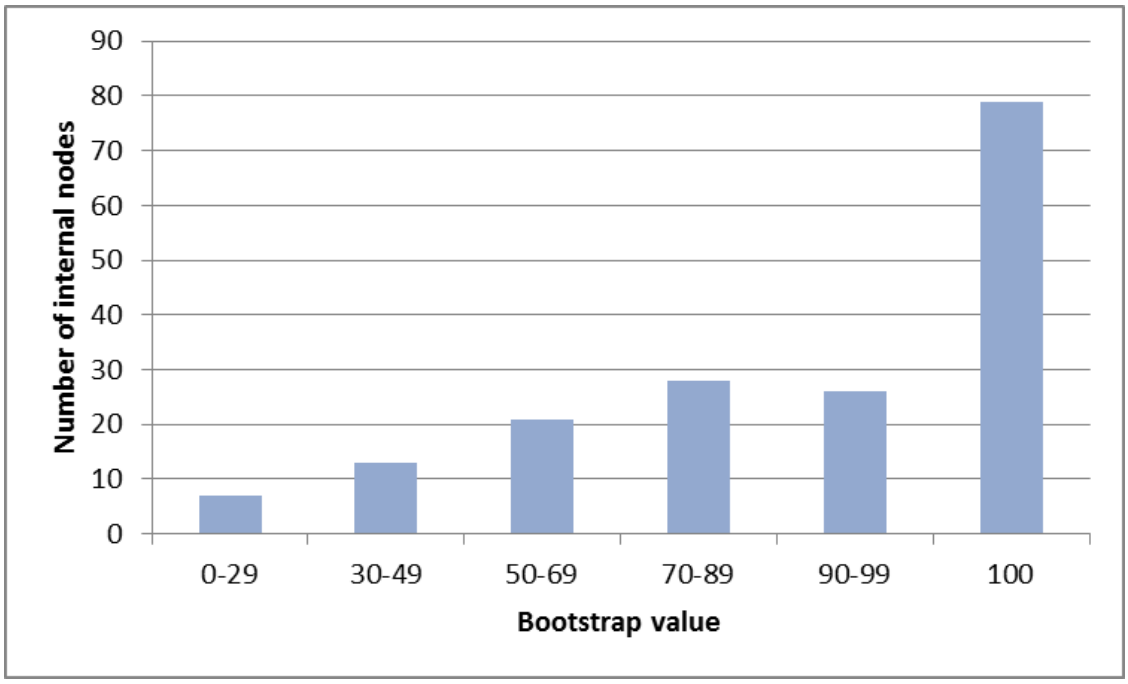


5.3.4 Phylogenetic analysis

5.3.4.1 Phylogenies using full genomes

Phylogenetic analysis using full genome sequences is well supported, with 76% (133/174) of internal nodes in the whole genome trees supported by bootstrap values ≥70 (**Figure 5.10**), which means the trees are very robust and reliable. Moreover the trees have defined clusters with distinct branching order.

Figure 5.10 Frequency histogram of internal node bootstrap support values in full genome maximum likelihood trees (all genotypes)



Phylogenetic analysis using maximum likelihood trees identified 19 clusters of related sequences, with 2–17 sequences per cluster (median 3) (**Table 5.2**, **Figure 5.11**, **Figure 5.12**, **Figure 5.13**, **Figure 5.14**, **Figure 5.15**, **Figure 5.16**, and **Figure 5.17**). Sequence clusters were defined as a group of monophyletic sequences with no discernible horizontal branch length; i.e. a group of sequences derived from a common evolutionary ancestor, or a single founder, not shared with any other group and with minimal divergence between them. It is these sequence clusters which are the basis of all further analysis.

Table 5.2 Sequence clusters identified by maximum likelihood phylogeny using full genome sequences. All GII.4 capsid genotypes are Sydney_2012. GII.4 polymerase genotypes are GII.Pe Sydney_2012 or GII.P4 New Orleans_2009

Cluster number	Genotype	Number of patients	Date range	Number of wards	Wards	Number of clinical specialties involved	Diversity within cluster †	Identified by infection control	Supported by classical epidemiology*
1	GI.P3_GI.3	4	15 days – 5 months	4	Koala, Robin, MOP, Bumblebee	3	17–56	No	No
2	GII.P2_GII.2	2	2 months	2	Penguin, PICU	2	7	No	No
3	GII.P7_GII.6	2	3 months	2	Starfish, Fox	2	17	No	No
4	GII.P7_GII.6	3	7 days	1	Bear	1	0	Yes	Yes
5	GII.P21_GII.3	17	3 months	6	Elephant, Lion, Safari, Butterfly, Rainforest, Bear	3	0–22	Partially	Yes (16/17)
6	GII.P21_GII.3	2	3 days	1	Safari	1	14	No	Yes
7	GII.P21_GII.3	6	1 month	3	Eagle, Robin, Safari	2	0–10	Partially	Yes
8	GII.P21_GII.3	2	2 months	1	Fox	1	11	No	Yes
9	GII.P21_GII.3	2	2 days	1	Fox	1	12	No	Yes
10	GII.P21_GII.3	9	10 months	2	Butterfly (8/9) and Bumblebee (1/9)	1	19–149	Partially	Yes
11	GII.Pe_GII.4	8**	2 months	2	Robin (7/8) and Eagle (1/8)	2	1–24	Partially	Yes
12	GII.Pe_GII.4	2	6 days	1	Rainforest	1	3	No	Yes
13	GII.Pe_GII.4	2	3 days	1	Fox	1	0	No	Yes
14	GII.Pe_GII.4	4	11 days	2	Giraffe (3/4), Safari (1/4)	1	1–4	Partially	Yes
15	GII.Pe_GII.4	3	3 days	1	Squirrel	1	1–3	Yes	Yes
16	GII.P4_GII.4	7	3 months	2	Lion (6/7), Butterfly (1/7)	1	0–35	Partially	Yes
17	GII.P4_GII.4	2	25 days	2	Butterfly, Bumblebee	1	14	No	Yes
18	GII.P4_GII.4	3	10 days	1	Rainforest	1	0–1	No	Yes
19	GII.P4_GII.4	2	19 days	1	Elephant	1	6	No	Yes

† expressed as the number of nucleotide differences across whole genome; * overlap in norovirus positive period and in-patient history; ** including one parent (NORO/51, father of NORO/52)

Figure 5.11 Maximum likelihood phylogeny of full genome sequences from norovirus episodes with a **GI.2** capsid genotype. Individual sequences are labelled with a unique patient identifier (NORO/XX) and specimen collection date

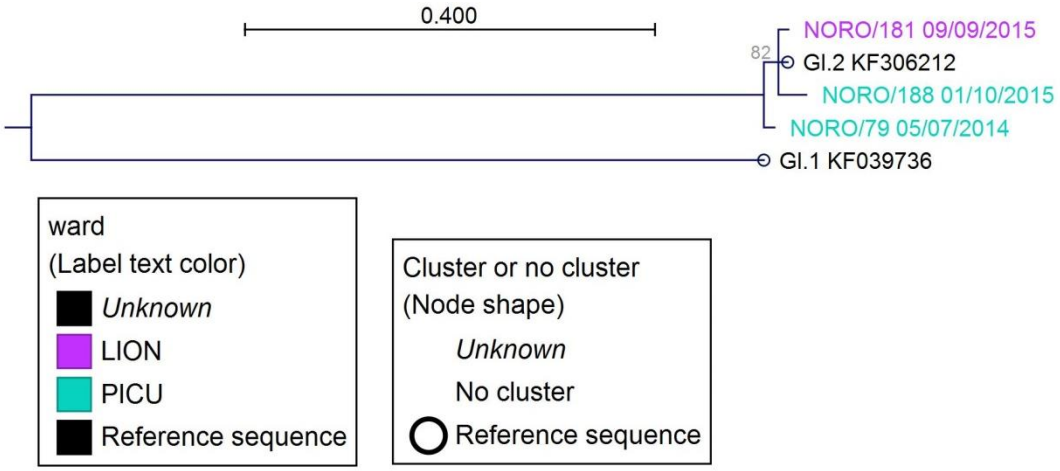


Figure 5.12 Maximum likelihood phylogeny of full genome sequences from norovirus episodes with a **GI.3** capsid genotype. Individual sequences are labelled with a unique patient identifier (NORO/XX) and specimen collection date

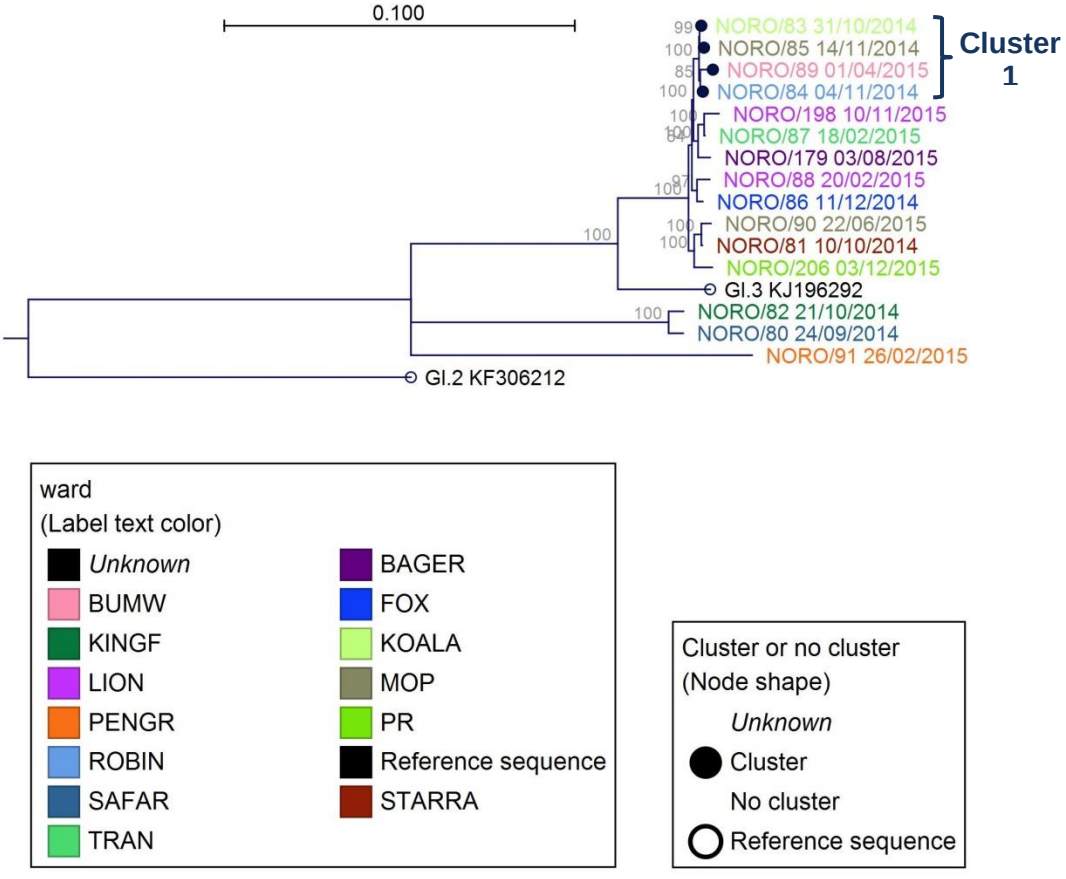


Figure 5.13 Maximum likelihood phylogeny of full genome sequences from norovirus episodes with a **GII.2** capsid genotype. Individual sequences are labelled with a unique patient identifier (NORO/XX) and specimen collection date

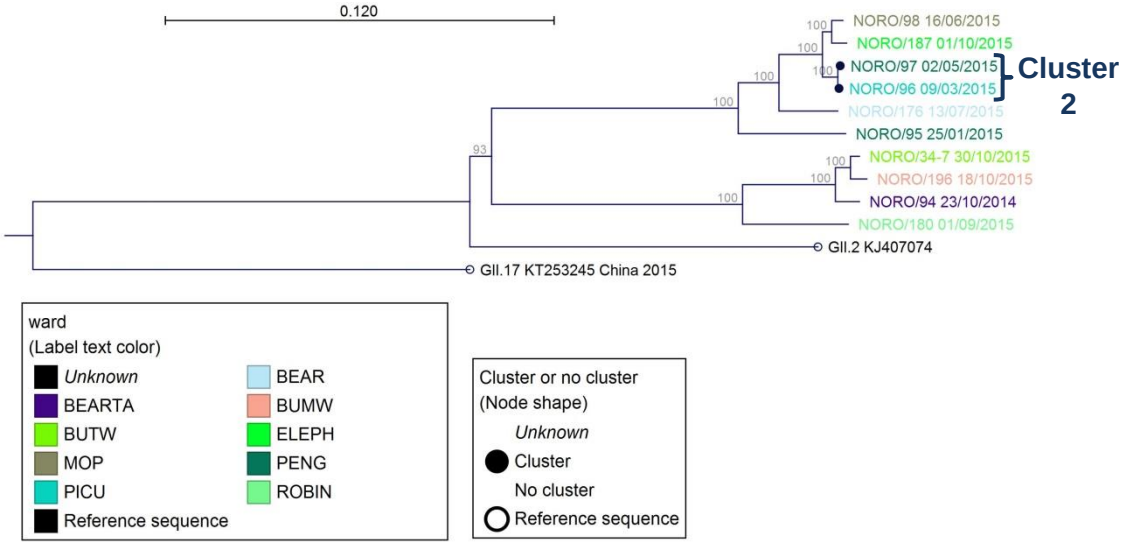


Figure 5.14 Maximum likelihood phylogeny of full genome sequences from norovirus episodes with a **GII.6** capsid genotype. Individual sequences are labelled with a unique patient identifier (NORO/XX) and specimen collection date

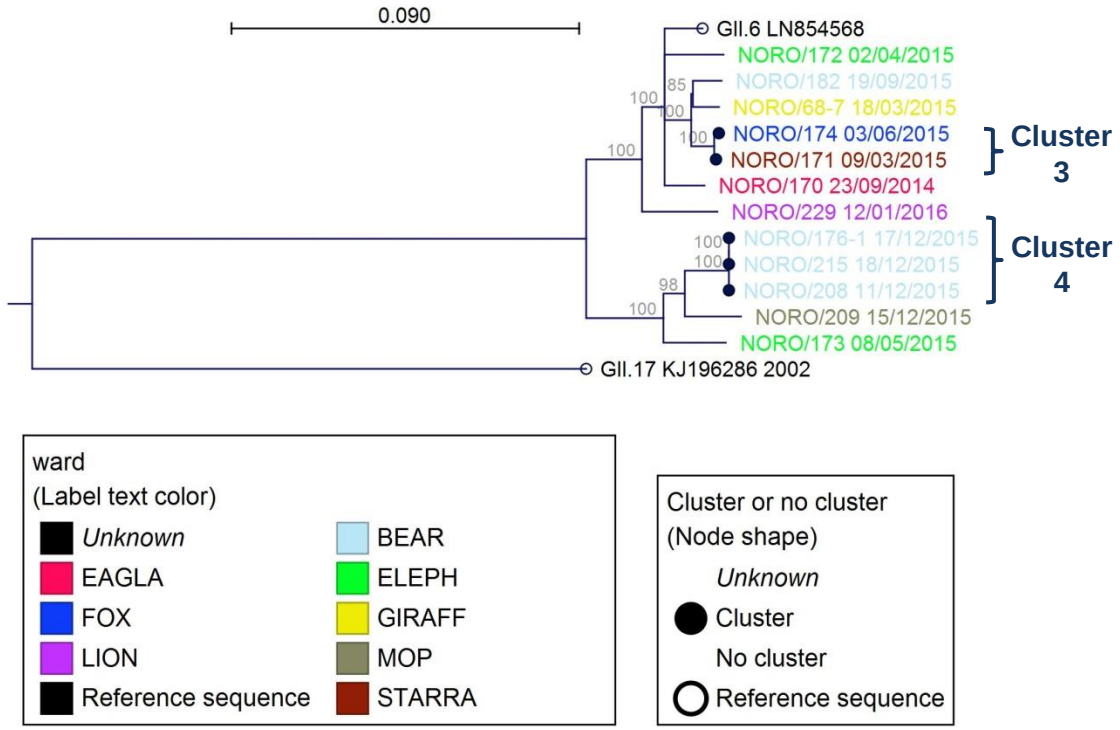


Figure 5.15 Maximum likelihood phylogeny of full genome sequences from norovirus episodes with a **GII.3** capsid genotype. Individual sequences are labelled with a unique patient identifier (NORO/XX) and specimen collection date

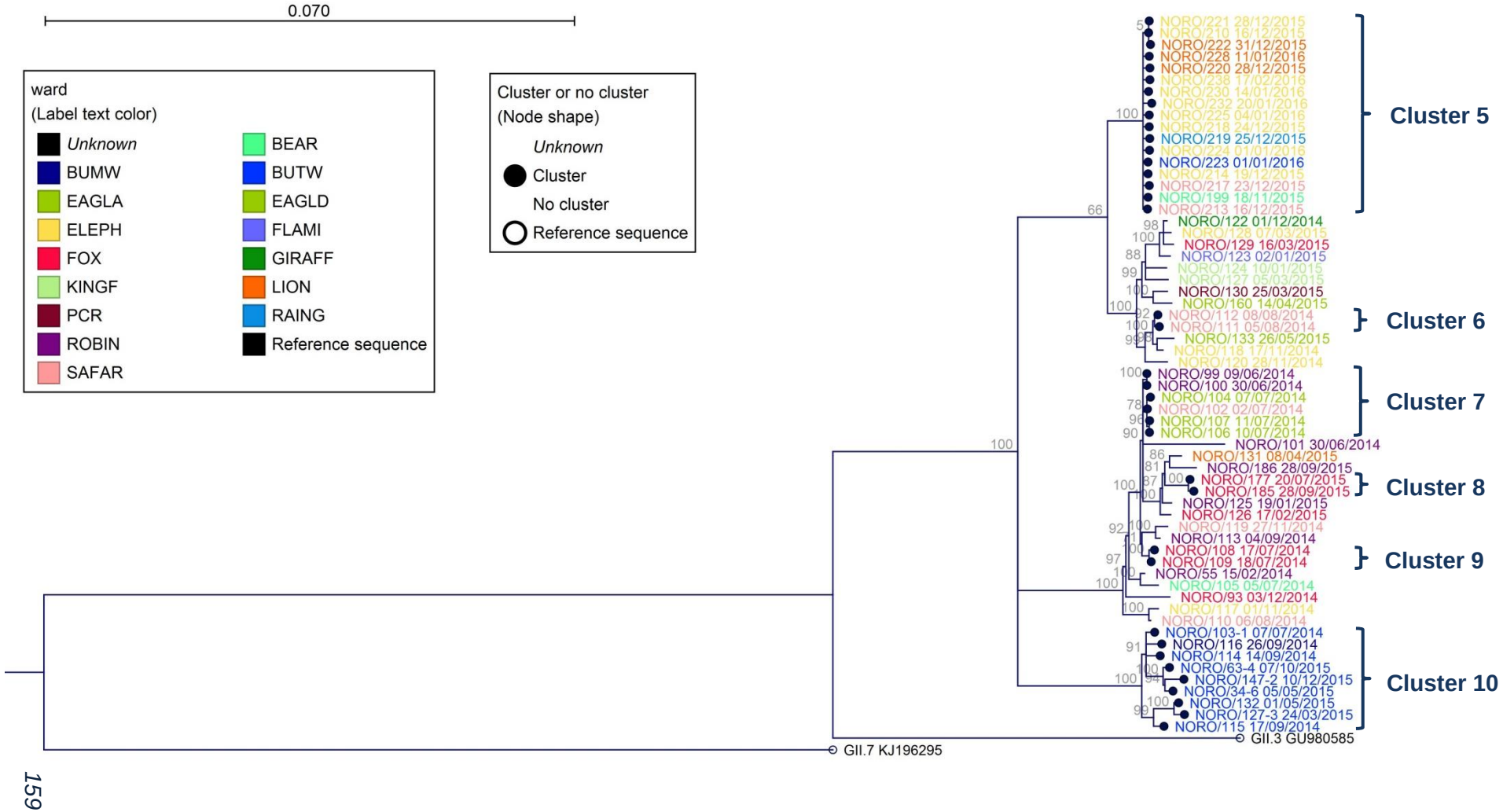


Figure 5.16 Maximum likelihood phylogeny of full genome sequences from norovirus episodes with a **GII.4** capsid genotype. Individual sequences are labelled with a unique patient identifier (NORO/XX) and specimen collection date

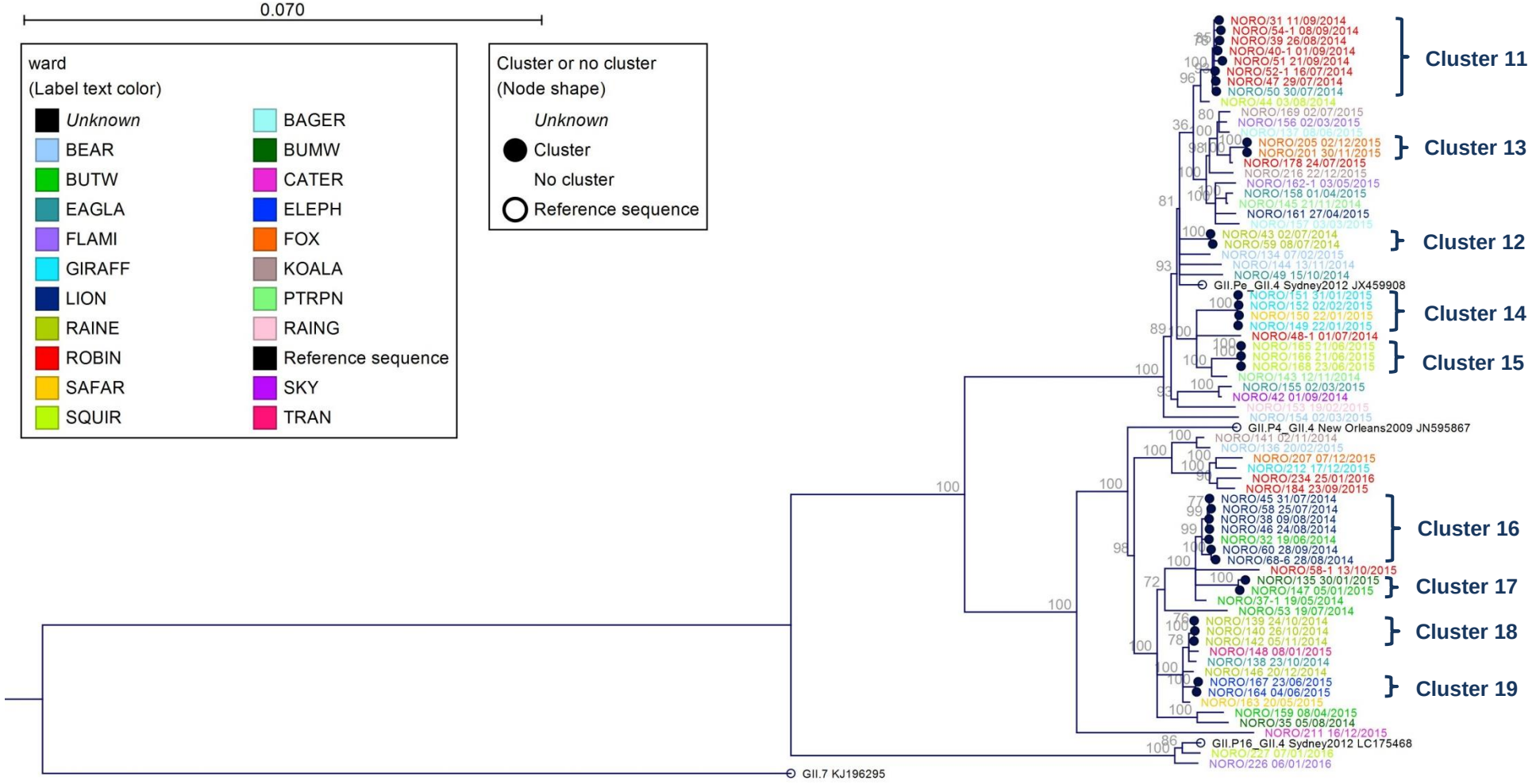


Figure 5.17 Maximum likelihood phylogeny of full genome sequences from norovirus episodes with a **GII.17** capsid genotype. Individual sequences are labelled with a unique patient identifier (NORO/XX) and specimen collection date

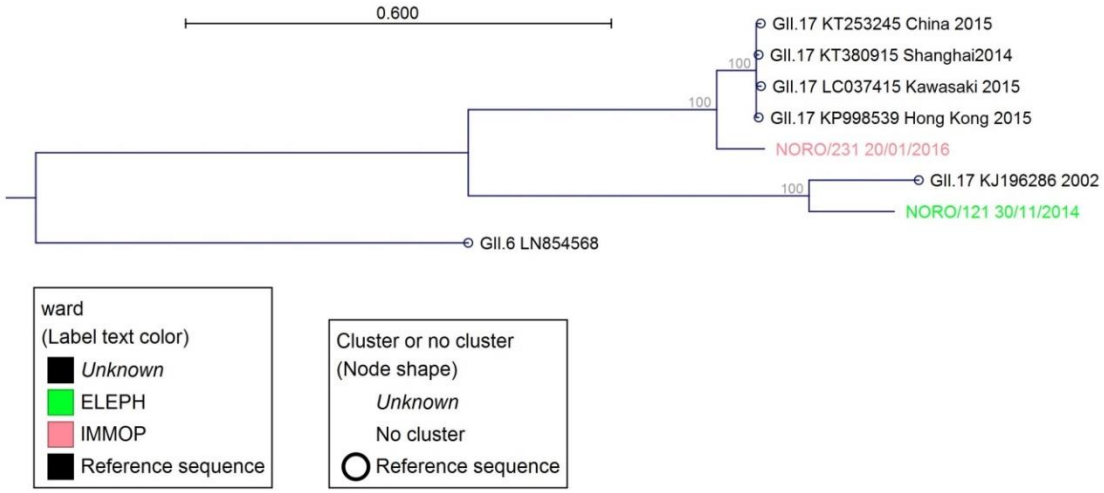
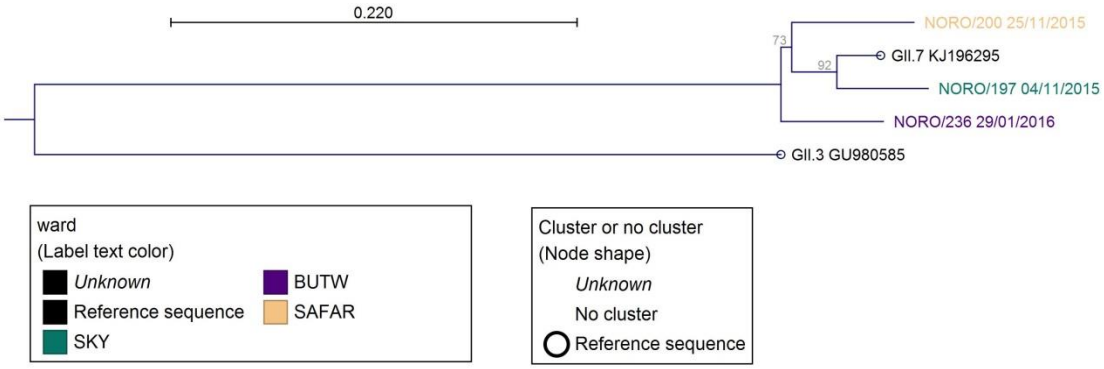


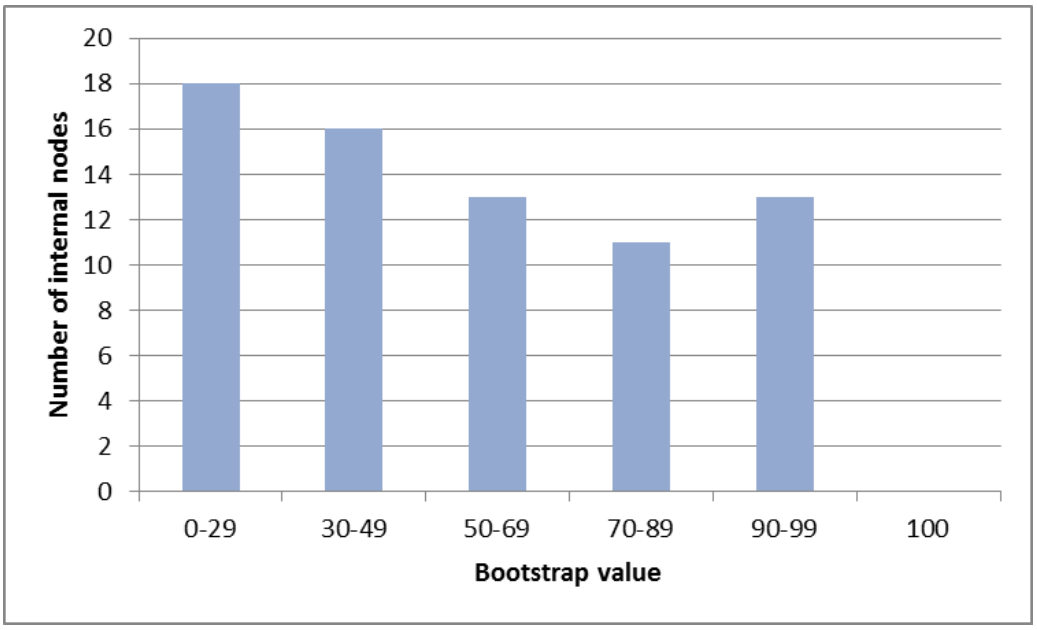
Figure 5.18 Maximum likelihood phylogeny of full genome sequences from norovirus episodes with a **GII.7** capsid genotype. Individual sequences are labelled with a unique patient identifier (NORO/XX) and specimen collection date



5.3.4.2 Phylogenies using P2 sequences

Maximum likelihood phylogeny of GII.4 episodes using the hyper-variable capsid P2 domain sequences (427 nt) generates a tree with very low bootstrap support; only 34% (24/71) of internal nodes are supported by bootstrap values ≥70 (**Figure 5.19**), which means the branching order, and therefore the tree, is neither robust nor reliable.

Figure 5.19 Frequency histogram of internal node bootstrap support values in P2 domain maximum likelihood trees (GII.4 norovirus only)



Since internal nodes with bootstrap values <70 are collapsed due to unreliability, the resulting tree has few clusters with the majority of sequences showing no branching order (**Figure 5.20**).

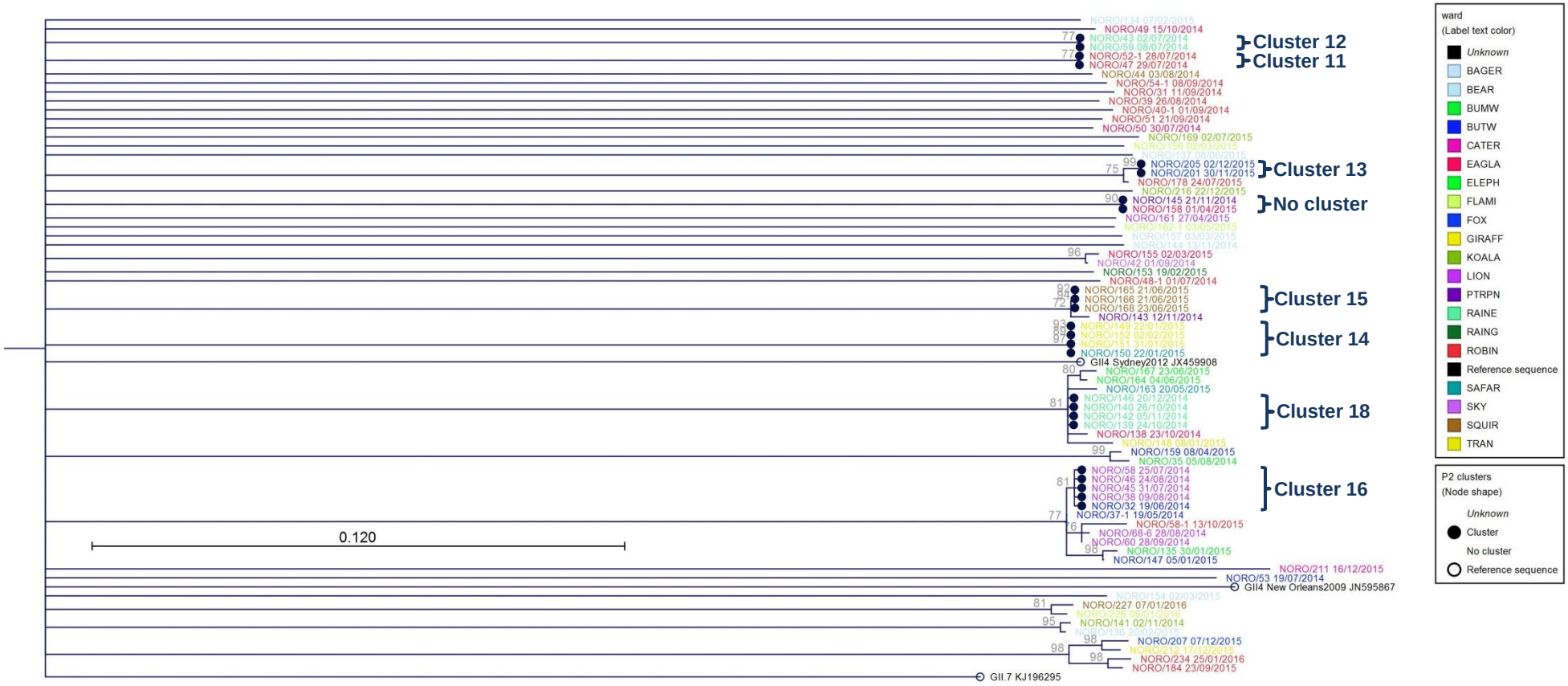
P2 sequences within clusters are not identical, with up to 6 nucleotide differences between sequences within a cluster. Only clusters with no SNPs are identifiable by P2 phylogeny (**Table 5.3**).

Table 5.3 Diversity in P2 region of GII.4 clusters identified by whole genome phylogeny. Cluster number refers to clusters identified using full genome sequences (in **Figure 5.16** on page 160)

Cluster number	Diversity in P2 region within cluster*	Identified by P2 phylogeny**
11	0—6	Partially (2/8 patients)
12	0	Yes
13	0	Yes
14	0	Yes
15	0	Yes
16	0—5	No
17	2	No
18	0	Yes
19	3	No

*Expressed as number of nucleotide differences across P2 region; **as shown in **Figure 5.20**

Figure 5.20 Maximum likelihood phylogeny of hyper-variable capsid P2 domain sequences for norovirus GII.4 episodes at GOSH. Cluster numbers refer to clusters identified using full genome sequences in Figure 5.16 on page 160



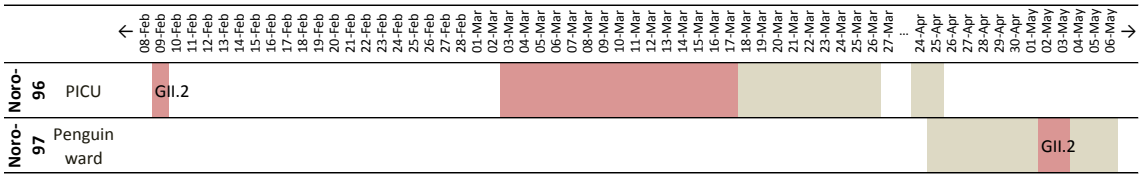
5.3.5 Molecular and classical epidemiology

5.3.5.1 Verification of sequence clusters

All of the clusters identified by phylogenetic analysis of full genome sequences were verified by looking at the in-patient history and duration of norovirus positivity of each patient to determine whether transmission between the patients in each cluster would have been possible.

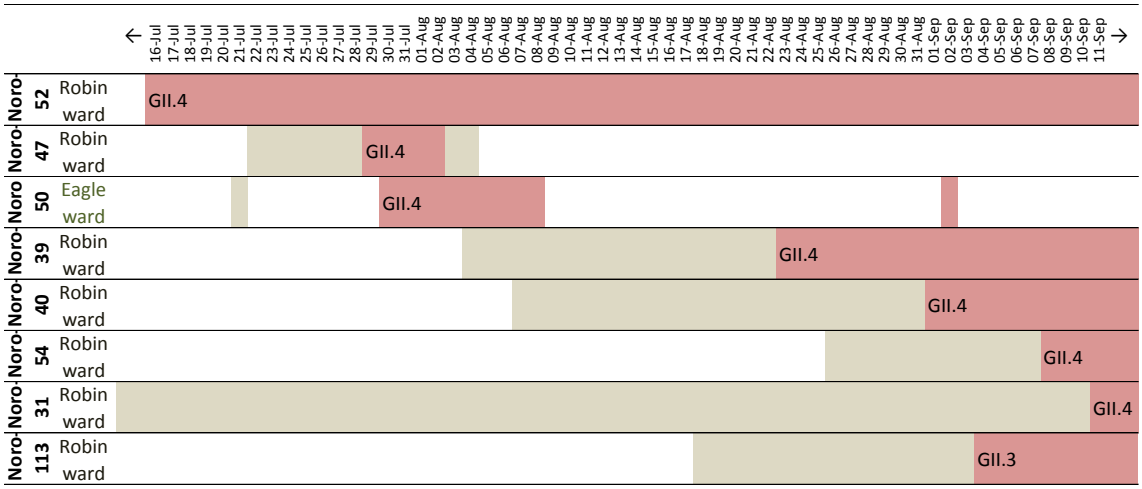
Three of the 19 clusters identified by maximum likelihood phylogeny, Cluster 1 (GI.P3_GI.3), 2 (GII.P2_GII.2) and 3 (GII.P7_GII.6), were not supported by classical epidemiology. In Cluster 1 (GI.P3_GI.3) all four patients were under the care of different clinical teams; three of the patients were admitted within 15 days of each other in 2014, the fourth five months later in 2015 from Portugal. All four patients were already norovirus positive on admission with no known contact between patients prior to admission; three patients had no prior outpatient appointments. In Cluster 2 (GII.P2_GII.2), the two patients were under the care of different clinical teams. There was less than one day overlap in admission dates between the two patients; the index patient was norovirus negative for six weeks prior to the second patient becoming infected (illustrated in **Figure 5.21**). In Cluster 3 (GII.P7_GII.6) the two patients were also under the care of different clinical teams. Both patients were norovirus positive on admission, with no known contact between them prior to admission; neither patient had ever been seen in outpatients. Cluster 1, 2 and 3 were excluded from further analysis.

Figure 5.21 Timeline of patients in Cluster 2. Classical epidemiology does not support the cluster as the first patient, NORO/96, was already norovirus-negative when the second patient, NORO/97, was admitted. NORO/97 has no prior outpatient history. Each was cared for by different clinical teams. Grey bars indicated the patient was an in-patient or outpatient and norovirus PCR negative; pink bars indicate the patient was an in-patient or out-patient and norovirus PCR positive



The remaining 16 clusters were supported by classical epidemiology, i.e. there was an overlap between patient's admissions whilst one of the patients was norovirus positive. As an example, this is illustrated for Cluster 11 in **Figure 5.22**.

Figure 5.22 Timeline of patients in Cluster 11, including an additional patient, Noro 113, who was norovirus positive on Robin ward at the same time as Cluster 11, but was infected with GII.P21_GII.3 norovirus. Classical epidemiology supports the cluster, assuming NORO/50 was infected at an outpatient appointment on 21st of July prior to admission. Grey bars indicated the patient was an in-patient or outpatient and norovirus PCR negative; pink bars indicate the patient was an in-patient or out-patient and was norovirus PCR positive



Five of the 16 clusters involved likely transmission at outpatient appointments. In two clusters (Cluster 14 and 16) the index cases attended hospital as an outpatient whilst the next patient in the cluster to be infected was an in-patient. Conversely in three clusters (Cluster 11, 7 and 5) secondary transmission occurred at outpatient appointments; 1/8, 3/6 and 1/17 of the patients in each cluster, respectively, attended outpatient appointments whilst other patients in the cluster were already in-patients and norovirus positive. Since outpatient clinics are located on a separate floor to in-patient wards, transmission in these patients must have occurred indirectly; for example via staff movement or contaminated fomites.

Classical epidemiology evidence does not support 1 of the 17 patients in Cluster 5; this patient was norovirus positive during the same time frame as the rest of the cluster, however he/she was a day-case admission, with no previous outpatient appointments or known contact with other patients in the cluster. For the purpose of on-going analysis this patient was excluded from the cluster.

Cluster 10 has the greatest nucleotide diversity compared to other clusters (19–149 compared to 0–35 SNPs) and spans the greatest time period (17 months compared to 2 days–3 months in other clusters), which calls into question the validity of the cluster. However the cluster is distinctly monophyletic and all patients but one (8/9) are on the same ward; the remaining patient is on an adjoining ward under the same clinical team.

Epidemiological evidence supports this cluster, with an overlap in admission dates and norovirus positivity establishing a plausible link between all patients.

5.3.5.2 Comparison of molecular epidemiology and classical Infection Control investigations

During the study period there were eight outbreaks of norovirus identified by the infection control team at GOSH, with 2–12 patients per outbreak (median 3) (**Table 5.4**).

Table 5.4 Norovirus outbreaks identified by the Infection Control team using classical epidemiology (time and place) with details of corresponding sequence cluster

Outbreak number	Wards	Number of identified patients	Sequence cluster	Extra patients in sequence cluster missed by infection control team
14-06 L	Butterfly	1 (plus 3 staff)	Cluster 16	N=6. All on Lion ward one month later (same CS)
14-07 E	Eagle	3	Cluster 7	N=3. Two Robin ward (different CS) 3 days and one month earlier; one outpatient (same CS).
14-09 B	Robin	8	Cluster 11	N=1. Eagle ward (different CS) same time.
14-09 H	Butterfly	5	Cluster 10	N=5. One Bumblebee ward (same CS) same time. Four same ward two and four months earlier and 2 months later.
15-01 O	Giraffe	2	Cluster 14	N=2. One on same ward, same time; one outpatient (same CS)
15-06 P	Squirrel	3	Cluster 15	-
15-12 H	Bear	3	Cluster 4	-
15-12 P	Elephant and Lion	12	Cluster 5	N=6. Two same ward same time; one Rainforest ward (same CS) same time; one Butterfly ward (same CS) same time; one outpatients (same CS) same time; one Bear ward (different CS) one month earlier.
TOTAL number of patients		37		
Median number of patients		3		

CS, clinical specialty; -, none missed

Four patients who were identified as being part of an infection control outbreak were not part of any sequence cluster (**Table 5.5**) therefore had been incorrectly assigned to an outbreak based on time and place (one patient in 14-09 B, one in in 14-09 H and two patients in 15-12 P). All four of these patients were infected with a different genotype compared to the other patients in that cluster.

Table 5.5 Comparison of norovirus transmission events identified by molecular epidemiology (this study) and classical infection control monitoring. Green indicates patients correctly identified by infection control monitoring; red indicates patients incorrectly identified by infection control monitoring

	Part of infection control outbreak	Not part of infection control outbreak	TOTAL
Transmission by molecular epi (part of verified sequence cluster)	**33**	**40**	73
No transmission by molecular epi (not part of sequence cluster)	**4**	**106**	110
TOTAL	37	146	183

Overall, conventional infection control monitoring missed 40/73 (55%) of linked infections. Assuming molecular epidemiology sequence clustering corroborated by evidence of patient contact to be the gold standard, it follows that the negative predictive value of classical infection control is 73% (106/146); i.e. only 73% of patients were correctly identified as not linked to other patients. On the other hand, the positive predictive value for classical epidemiology (i.e. proportion of correctly identified linked cases) is high at 89% (33/37); if two patients are identified as linked by classical infection control investigations then this is likely to be true. This means that in this setting classical infection control investigations lack sensitivity (missing patients who are linked) but are specific (do not link patients who are not part of an outbreak).

5.3.5.3 Clusters fully or partially predicted by Infection Control

Two of the eight outbreaks identified by Infection Control correctly identified all patients involved in the transmission cluster.

However for six of the eight outbreaks, extra linked cases were identified by molecular epidemiology that were missed by the Infection Control team, totalling 23 extra patients. The majority of these were norovirus infections that occurred at a similar time on different wards with a shared clinical specialty (7/23) or at a different time, either on the same ward or with shared clinical specialty (4/23 and 6/23, respectively) (**Table 5.6**).

Table 5.6 Extra patients in sequence clusters that were partially recognised by Infection Control investigations (n=23 patients)

	Same time	Different time
Same ward	3	4
Different ward or outpatient but same CS	7	6
Different ward and different CS	1	2

CS, clinical specialty

5.3.5.4 Clusters missed by infection control

There are eight sequence clusters that were missed completely by the infection control team (**Table 5.2** on page 156), totalling 17 patients. The majority (11/17) of these patients were infected on the same ward at a similar time; however each cluster only involved two or three patients. In addition, sequence clusters were not recognised as outbreaks by Infection Control if patients were on the same ward but separated by time (4/17) (**Table 5.7**).

Table 5.7 Patients in sequence clusters that were not recognised by Infection Control investigations (n=17 patients)

	Same time	Different time
Same ward	11	4
Different ward or outpatient but same CS	0	2
Different ward and different CS	0	0

CS, clinical specialty

5.3.6 Transmission dynamics at GOSH

Overall, 73/183 (40%) of sequences from patients with norovirus at GOSH are part of a sequence cluster, suggesting they are involved in a transmission chain. Assuming that each of the 16 sequence clusters includes a source patient who, whilst transmitting to other patients were themselves not infected by another patient, only 57/183 (31%) of patients during this study period were infected by another patient (**Table 5.8**); 77% (44/57) of these patients were immunocompromised and 23% (13/57) were medical or surgical patients. Conversely 111/183 (60%) of cases were not linked to any others during the 18 month study period; 50% (55/111) of these patients were immunocompromised and 50% (56/111) were medical or surgical patients. A quarter of all patients (44/183, 24%) acquired a norovirus infection whilst an in-patient at GOSH during the study period but were not part of any sequence cluster or they were the index patient therefore were infected from an unknown source; 59% (26/44) of these were immunocompromised and 41% (18/44) were medical or surgical patients.

Table 5.8 Summary of sources of infection at GOSH during study period, July 2014–February 2016

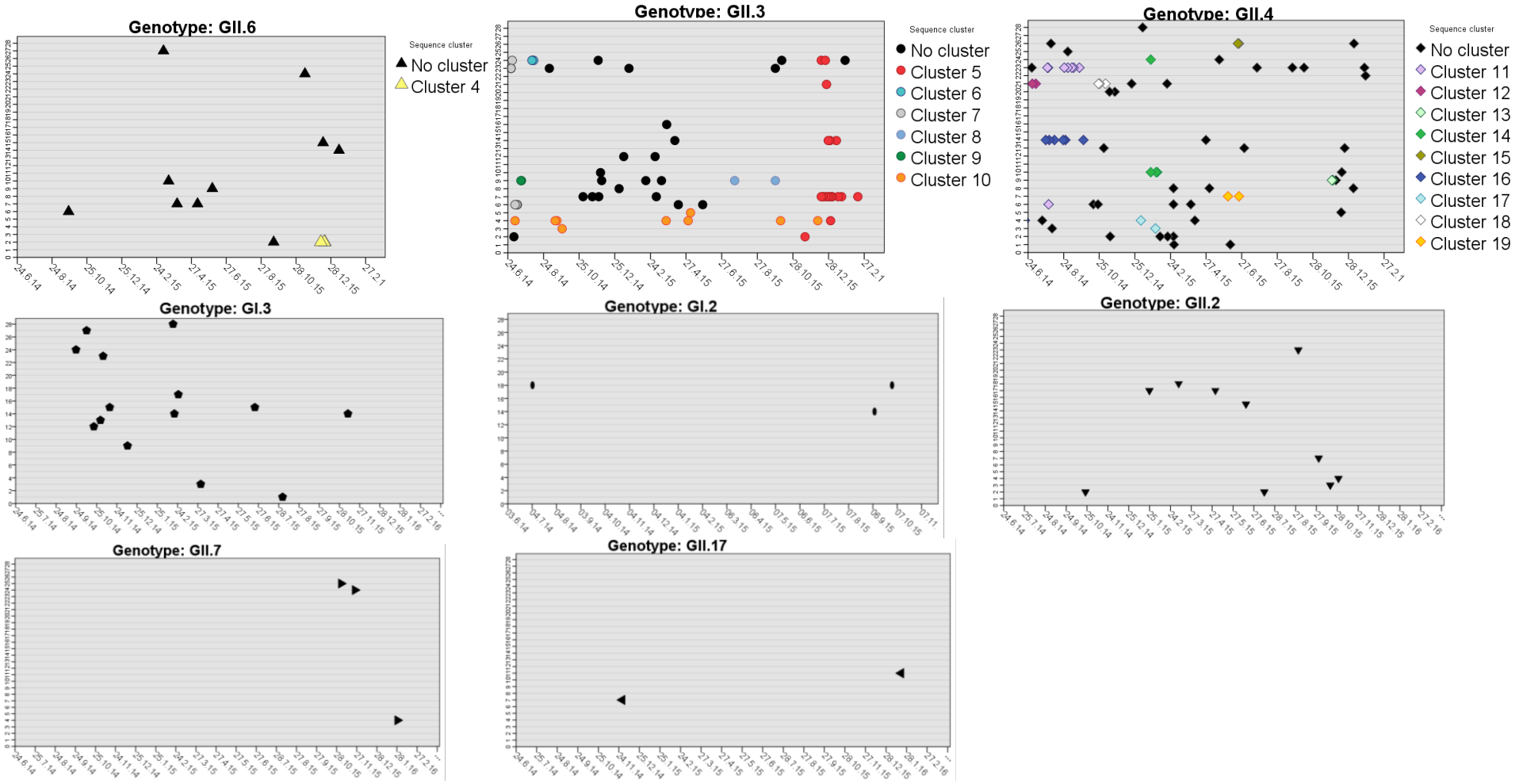
Source of infection	Number of patients (n=183)	Proportion of patients immunocompromised
Another patient (part of a transmission cluster)	57 (31%)	44/57 (77%)
Outside of the hospital (norovirus positive on admission)	82 (45%)	39/82 (48%)
Unknown (hospital acquired infection but not transmission from another patient)	44 (24%)	26/44 (59%)

Only 4/16 clusters had an index patient who was norovirus positive on admission (POA) with a source of infection outside the hospital; in the remaining 12/16 clusters (75%) the index patient had a hospital acquired infection for which the source of infection is unknown.

With the exception of one GII.6 cluster involving two patients, all transmission events involved norovirus genotypes GII.3 or GII.4. Sequence clustering from maximum likelihood phylogenies identified clusters of transmission within genotypes and, conversely, excluded transmission between patients infected with the same genotype (**Figure 5.23**).

Of the eight instances where norovirus occurred on the same ward, at the same time and caused by the same GII.4 genotype (circled in **Figure 5.9** on page 154), 7/8 were confirmed as linked by phylogenetic analysis. Similarly, 7/8 instances involving GII.P21_GII.3 infection on the same ward at the same time were confirmed as linked by phylogenetic analysis.

Figure 5.23 Distribution of norovirus episodes across wards, July 2014–February 2016, separated by genotype and colour coded by sequence cluster (identified by full genome maximum likelihood phylogeny). Black points indicate no cluster. Each line on the y axis represents a different ward. X axis indicates the date of infection.



5.3.7 Molecular epidemiology in a non-paediatric general hospital (NNUH)

Of the 62 norovirus stool samples sequenced from NNUH, 60 (97%) were GII.4 Sydney_2012 strains; 12/60 (20%) were GII.Pe_GII.4 (Sydney_2012 ORF1) and 48/60 (80%) were GII.P4_GII.4 (New Orleans_2009 ORF1).

Maximum likelihood phylogeny of the GII.4 infections using whole genome sequences shows that norovirus infections at NNUH fall into four clusters, labelled A–D in **Figure 5.24**. The majority of GII.4 infections (46/60) are part of Cluster A which is a GII.P4 New Orleans_2009_GII.4 Sydney_2012 strain. In total, 53/60 (88%) of patients were infected from another patient. Diversity within each cluster is very limited, with a maximum of 6 SNPs across the whole genome within a single cluster (**Table 5.9**).

Phylogenetic analysis using the P2 domain sequences from NNUH correctly identifies all of the clusters seen using full genome sequences (**Figure 5.25**). All of the sequences within a cluster have identical P2 sequences (0 SNPs), with the exception of one sequence pair which has 1 SNP (**Table 5.9**).

Table 5.9 NNUH sequence clusters identified by whole genome and P2 maximum likelihood phylogeny. All clusters were identified by both analyses

Cluster	Number of sequences	Date range	Within cluster diversity in full genome*	Within cluster diversity in P2 region*
A	46	1 week	0–6†	0–1
B	2	2 weeks	0	0
C	2	2 days	9	0
D	7	10 days	0–2	0

*expressed as number of nucleotide differences; †up to 34 if include NOR_2191 (all SNPs occur in 100 nt region at 5’ end)

Figure 5.24 Maximum likelihood phylogeny of NNUH GII.4 **full genome** sequences. Sequences are labelled with the specimen collection date

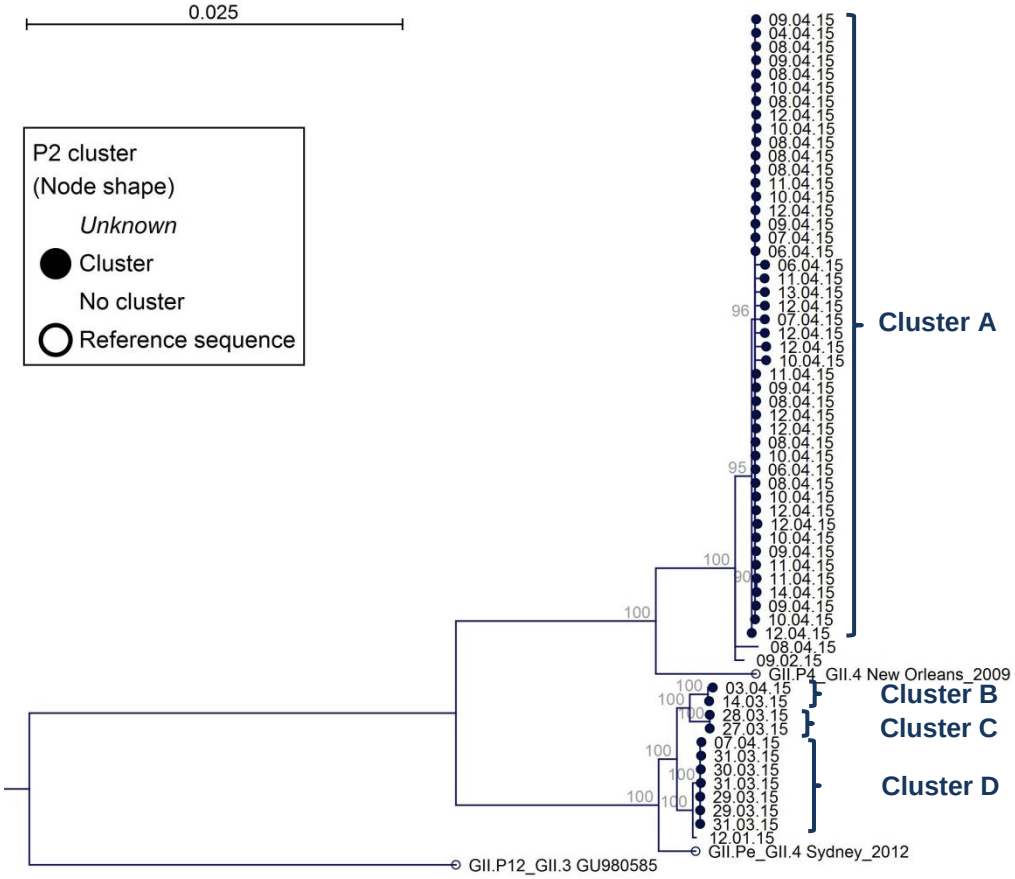
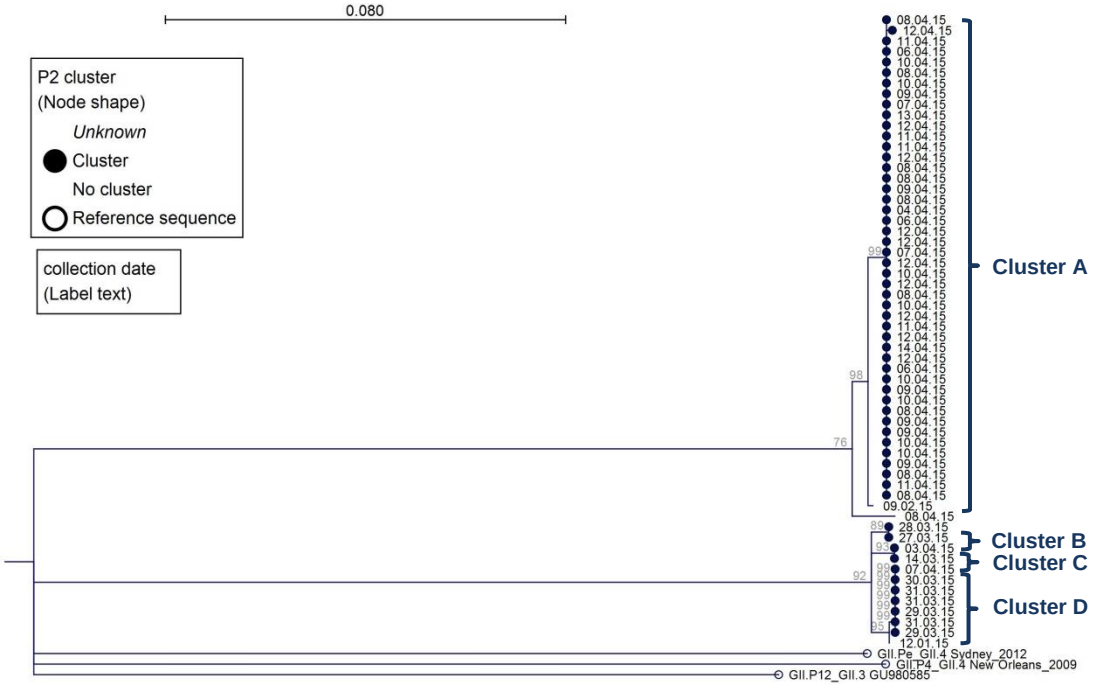


Figure 5.25 Maximum likelihood phylogeny of NNUH GII.4 **P2 domain** sequences. Sequences are labelled with the specimen collection date



5.3.8 Persistence vs. re-infection in longitudinally sampled patients

5.3.8.1 Continuously positive patients

Eighteen of the 25 longitudinally sampled patients were continuously norovirus PCR positive in between collected samples (median 129 days from first to last sequenced sample, range 7–466) (**Table 5.10**). 3/16 (19%) of these patients were infected with a different genotype by the last sample, including two (Px 101 and Px 65) who were co-infected with two genotypes. Two of the patients who were re-infected with a different genotype were chronically infected for a long time (321 and 466 days), however for one patient only 58 days had passed. The remaining 15/18 patients (83%) remained infected with the same genotype. All longitudinal samples which remained the same genotype cluster together by patient in the phylogenetic tree (**Figure 5.26**). This suggests on-going infection is caused by persistence with the same virus, rather than re-infection with a new virus of the same genotype.

The median duration of infection was 322 days (range 58–738 days) in patients who had a super-infection and 215 days (range 14–711 days) in those who did not. The difference in duration of infections was not statistically significant (P = 0.360).

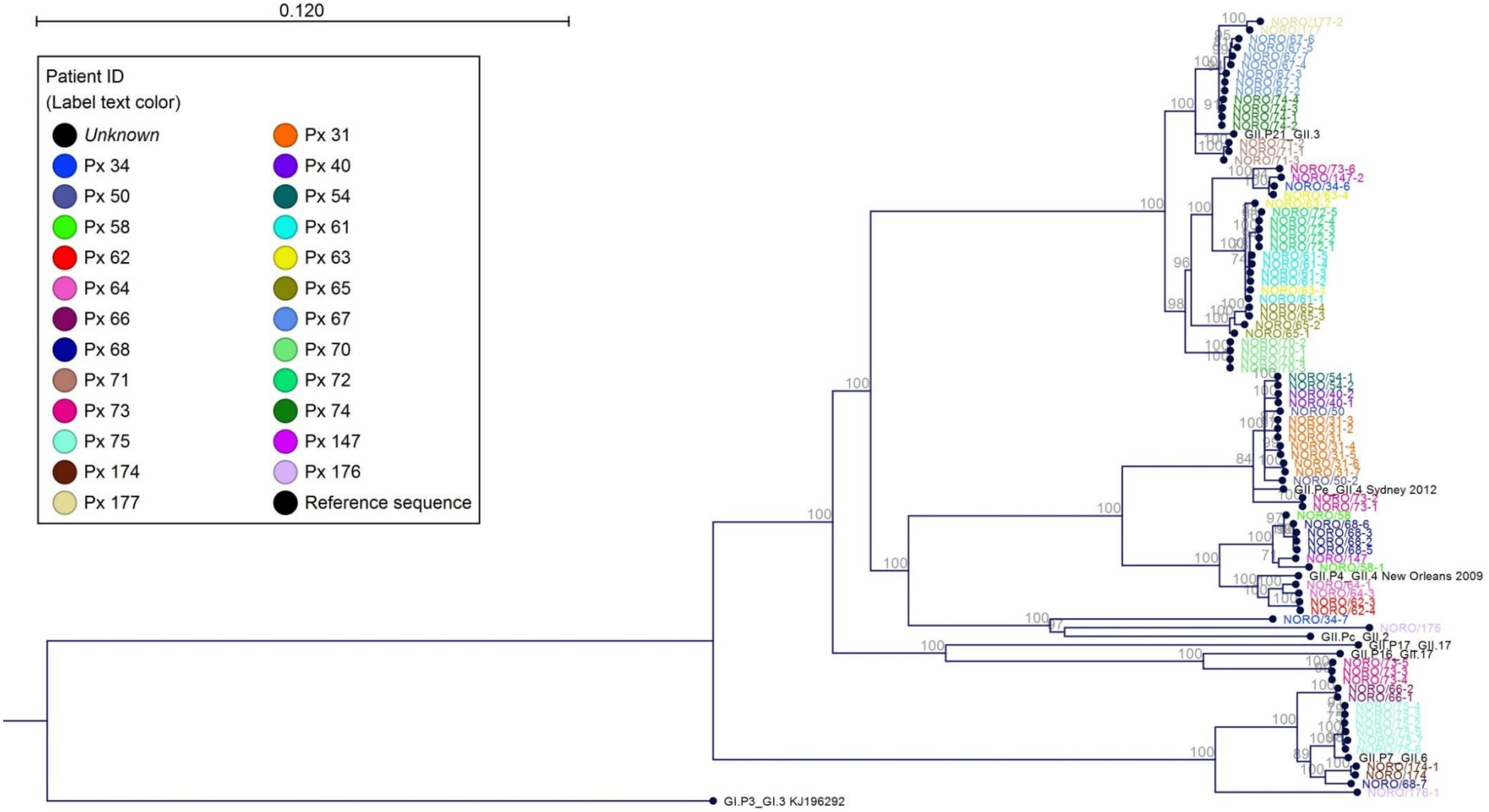
5.3.8.2 Chronically infected patients with undetected period

Nine of the 25 longitudinally sampled patients had a period of norovirus PCR negative stools in between the first and last sequenced samples (median 153 days undetected, range 9–466) (**Table 5.10**). 5/9 patients (56%) were infected with a different genotype after the negative period. 4/9 (44%) remained infected with the same genotype. One of the patients who remained infected with the same genotype (Px 63) is identified as a new infection based on phylogenetic analysis (**Figure 5.26**), in which the latest sample does not cluster with earlier samples. Those who remained infected with the same genotype (3/9, 33%) had the shortest interval during which norovirus was not detected; less than two months compared to 5–15 months for those who were re-infected.

Table 5.10 Summary of longitudinally sampled norovirus infections

Patient ID	Number of samples	Days between first and last sample	Negative in between samples	Genotype	Re-infection (Yes/No)
63	3	1006	No (135 days)	GII.P21_GII.3	No
			Yes (466 days)	GII.P21_GII.3	Yes
101	2	466	No	GII.P21_GII.3 to mixed GII.P21_GII.3/ GII.P7_GII.6	Yes
50	2	446	No	GII.Pe_GII.4	No
58	2	445	No	GII.P4_GII.4	No
65	5	321	No	GII.P21_GII.3 to mixed GII.P21_GII.3/GII.Pe_GII.4	Yes
64	2	243	No	GII.P4_GII.4	No
67	7	197	No	GII.P21_GII.3	No
177	2	190	No	GII.P21_GII.3	No
174	2	163	No	GII.P7_GII.6	No
61	5	123	No	GII.P21_GII.3	No
66	2	35	No	GII.P7_GII.6	No
74	4	28	No	GII.P21_GII.3	No
71	3	14	No	GII.P21_GII.3	No
70	4	13	No	GII.P21_GII.3	No
62	2	7	No	GII.P4_GII.4	No
54	2	7	No	GII.Pe_GII.4	No
40	2	7	No	GII.Pe_GII.4	No
147	2	339	Yes (11 months)	GII.P4_GII.4 to GII.P21_GII.3	Yes
68	5	202	Yes (6 months)	GII.P4_GII.4 to GII.P7_GII.6	Yes
34	2	178	Yes (5 months)	GII.P21_GII.3 to GII.Pe_GII.2	Yes
176	2	157	Yes (5 months)	GII.P2_GII.2 to GII.P7_GII.6	Yes
73	6	152	No (58 days)	GII.Pe_GII.4 to GII.P16_GII.17	Yes
			Yes (57 days)	GII.P16_GII.17 to GII.P21_GII.3	Yes
31	7	123	Yes (56 days)	GII.Pe_GII.4	No
72	5	73	Yes (23 days)	GII.P21_GII.3	No
75	6	62	Yes (9 days)	GII.P7_GII.6	No

Figure 5.26 Full genome maximum likelihood phylogeny of longitudinal norovirus sequences. Px, patient number



5.4 DISCUSSION

5.4.1 Full genomes for norovirus genotyping

In Chapter 3 PCR and Sanger sequencing of the capsid shell domain identified the circulating genotypes at GOSH over a 12 month period (July 2014–July 2015) and showed a broad range of circulating genotypes. This study uses full genome sequences to determine the norovirus genotype, thus identifying both the polymerase and capsid genotypes over a 19 month period (July 2014 – February 2016). The breadth of genotypes is replicated, with a third (34%) of sequences identified as GII.P21_GII.3 and a third (27%) identified as non GII.4 genotypes. 39% of sequences are identified as having a GII.4 Sydney_2012 capsid sequence, however only half (55%) of these sequences have a GII.Pe Sydney_2012 polymerase; with the remaining 41% a GII.P4 New Orleans_2009 polymerase and 4% a GII.P16 polymerase.

The Sydney_2012 norovirus variant emerged globally in the winter of 2012/2013, causing an early norovirus season in the UK between October and December 2012, instead of the usual season of January to March. By the 2013/2014 season Sydney_2012 had replaced the previously circulating New Orleans_2009 variant, however both variants were in circulation during the 2012/13 winter season. Our study shows that at GOSH 45% of circulating GII.4 norovirus episodes during 2014–2016 are recombinant strains with non-Sydney_2012 polymerases. The same is seen at Norfolk and Norwich, in which 77% of GII.4 sequences have the New Orleans_2009 polymerase (GII.P4_GII.4), confirming that recombinant GII.4 genotypes are not unique to our population at GOSH.

This New Orleans_2009/Sydney_2012 GII.4 recombinant strain has previously been described in the literature using PCR to sequence partial or full genomes however all, but one, of these reports are from the 2012/2013 winter norovirus season [147, 192, 196, 197] during which the two strains were co-circulating. There has been only one report of this recombinant strain since, by Bruggink *et al.* in 2016 [198]. Bruggink *et al.* report that this strain was in circulation at low frequency in Victoria, Australia in 2015, causing 3/64 (5%) reported outbreaks in the region; in 2016 it was reportedly responsible for the majority (17/23, 74%) of their outbreaks. Bruggink *et al.* suggest this may be a novel epidemic strain, however this suggestion is disputed by Fonager *et al.* [199] who oppose the novelty of the strain since it was detected in 2012/13. The data presented in this chapter supports Fonager *et al.*'s stance since at GOSH the New Orleans_2009/Sydney_2012 recombinant was already in circulation, at high frequency,

in 2014. It is likely that these recombinant genotypes have been in circulation since the 2012/2013 season however, due to a tendency to capsid genotyping alone, they have not previously been recognised.

Norovirus infections caused by GII.P16_GII.4, causing three infections in this study, were until very recently unknown. In 2016, nine infections in an outbreak in Kawasaki, Japan, were also caused by GII.P16_GII.4 [200]. There are currently no other reports of this recombinant genotype therefore it is not known how widespread or common it is, however given that it has now been detected in Japan and the UK it is likely to have a global distribution. Despite sampling for this study spanning almost two years, it was not detected in our population until January 2016 which suggests it could be a newly emerging genotype.

Genotyping using the full genome has revealed co-circulation of multiple GII.4 lineages, which provides potential insight into the global epidemiology of norovirus. However genotyping using the full genome is also useful at a local level; determining the polymerase and capsid genotype has allowed identification of multiple GI.3, GII.2, GII.4 and GII.17 lineages. Capsid typing alone would have falsely assumed circulation of a single lineage for each genotype. A combination of the polymerase and capsid genotype provides extra information to exclude the occurrence of a transmission event.

5.4.2 Full genomes vs. P2 region for outbreak investigation

Whilst in immunocompetent patients the P2 region between linked patients is expected to be identical [140, 143]*, in vivo* evolution of norovirus in immunocompromised hosts introduces single nucleotide polymorphisms (SNPs) between the P2 region of linked patients [144]. Consequently, as demonstrated in this chapter, in the context of an immunocompromised population the P2 region is insufficient for outbreak investigations. The short sequence (427 nt) combined with SNPs occurring between linked patients, results in phylogenetic trees with very poor bootstrap support and very limited ability to detect transmission clusters. Consequently in the context of an immunocompromised population using only the P2 sequence for phylogenetic analysis is not informative.

Conversely in the context of an immunocompetent population such as at NNUH, in which limited or no viral evolution is expected to occur prior to transmission, sequencing the P2 region alone appears to be sufficient to identify transmission clusters. However there are limitations in comparing the two datasets to draw conclusions about immunocompetent and immunocompromised populations.

The NNUH population was chosen for comparison to GOSH as it represents a largely immunocompetent population, compared to GOSH's largely immunocompromised, for which specimens were collected for outbreak investigation. However there are other differences between the two populations which could account for the differences seen in norovirus molecular epidemiology. The first is that NNUH patients are adults (median sampled age 84 years, range 20–97) whereas GOSH patients are children (median sampled age 2 years, range 5 weeks to 16 years). This may explain the difference in breadth of genotypes between the two populations, with 3% of infections caused by non-GII.4 norovirus at NNUH compared to 60% at GOSH during the same sampling period. Additionally, at GOSH patients are screened for norovirus on admission whereas at NNUH they are not, which may also have resulted in a greater breadth of genotypes at GOSH since patients with community acquired norovirus infections are reported to have a higher proportion of non-GII.4 than those which are hospital acquired [130].

At GOSH 60% of the beds are in isolated rooms, whereas at NNUH the majority of beds are on open wards. It has previously been shown that norovirus outbreaks are poorly controlled in hospitals with nightingale style wards compared to isolated rooms [96] therefore this is likely to account for the difference in cluster sizes between the two populations (largest cluster in GOSH was 17 patients, compared to 46 patients at NNUH).

Lastly a confounding factor which would account for the difference in within cluster diversity seen between GOSH and NNUH populations is that whilst the GOSH population was sampled over a 19 month period, the NNUH population was predominantly sampled over 2 months. It is possible that over a longer sampling period the molecular epidemiology at NNUH could reveal clusters of norovirus infections with greater diversity than is seen in this study; consequently a more extensive sampling period, encompassing at least one calendar year to account for changes in norovirus seasonality, should be performed to confirm the findings observed in this study.

Full genome sequencing is critical to identify transmission in an immunocompromised population; full genomes allow phylogenetic reconstruction with very high bootstrap support, and thus reliable branching order, despite SNPs occurring between linked sequences. This allows reliable identification of sequence clusters. Whilst the full genome sequence may not be required for phylogenetic analysis in an immunocompetent population, it will nonetheless add value by identifying the polymerase genotype of the infecting virus.

5.4.3 Classical and molecular epidemiology to identify transmission events

All of the outbreaks identified by the infection control team, based on the occurrence of multiple norovirus episodes connected by time and place, were corroborated by molecular epidemiology using full genome sequences. The only exception to this was four patients who were included in outbreaks but were shown to be infected with a different genotype, therefore were not actually linked to the outbreak. Nevertheless in these instances transmission was occurring between other patients on the same ward therefore infection control efforts were not wasted by misidentifying these patients.

Conversely, many instances of transmission were missed by the Infection Control team. These largely fall into two categories. First, patients part of a recognised outbreak but not identified because they are on a different ward, albeit with shared clinical teams. Second, patients in clusters that occur on the same ward, but are completely missed by Infection Control because only two patients are involved.

The high PPV (89%) of classical epidemiology investigations suggests that cases linked by classical Infection Control monitoring are likely to be truly linked based on full genome sequences. However, the low NPV (73%) suggests that it is possible that cases that are not linked by infection control monitoring investigations are in fact linked by molecular epidemiology. The implication of this is that the true extent of an outbreak may not be recognised and therefore the outbreak not brought under control.

Phylogenies should not be interpreted in isolation; to determine whether the sequencing data represents transmission between patients it is necessary to determine whether there has been any contact between patients. In three of the nineteen clusters identified by full genome phylogeny there was no plausible link between patients. There is a possibility that there has been unidentified contact between patients, either via undocumented hospital attendance at GOSH, contact at another hospital or in the community. However in Cluster 1 in particular this is unlikely since one of the patients was admitted from Portugal. It may be that unrelated sequences in these genotypes (GI.P3_GI.3 for Cluster 1, GII.P2_GII.2 Cluster 2 and GII.P7_GII.6 for Cluster 3) are less variable compared to unrelated sequences in GII.3 and GII.4 infections; this will make phylogenetic analysis for outbreak investigation less useful where these genotypes are concerned as similar sequences do not necessarily indicate transmission. Extensive analysis of sequence diversity between unrelated sequences of the same genotype will aid with interpretation of phylogenies in this scenario,

however there is currently only one publically available full genome sequence for GI.P3_GI.3 and one for GII.P2_GII.2 therefore at present this analysis is not possible.

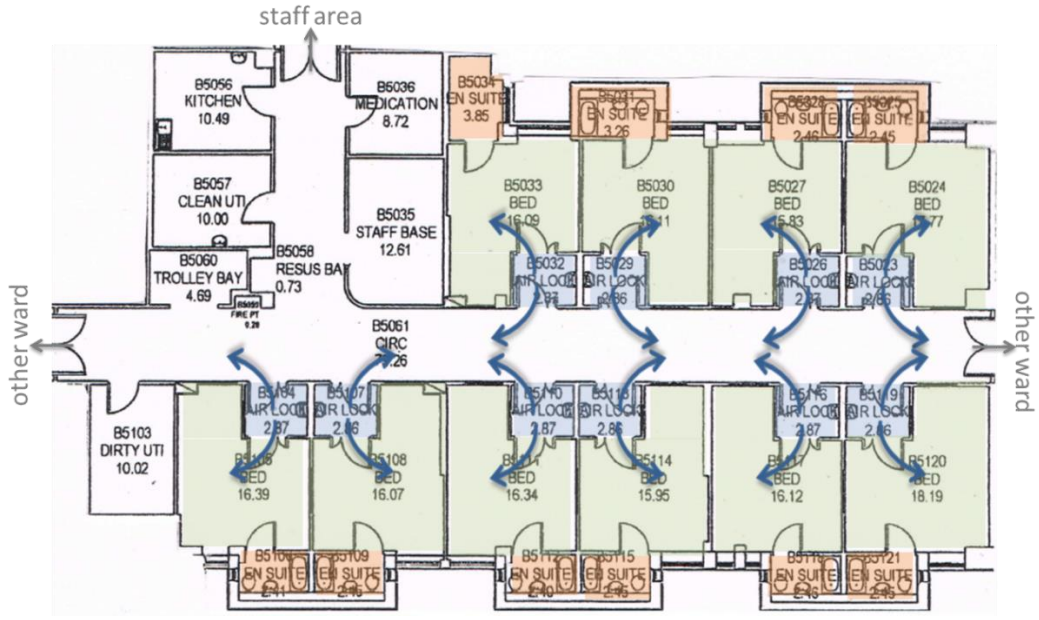
5.4.4 Indirect transmission of norovirus

Data presented in this chapter suggests that out-patients are both a source of infection and at risk of infection from in-patients; 5/16 clusters included an out-patient who visited hospital but was not admitted. This corroborates earlier findings by Sukhrie *et al.* [140] who demonstrated that out-patients were involved in 5/14 clusters identified with P2 sequence phylogenies.

These out-patients did not come into direct contact with the other patients in the cluster therefore transmission must have occurred indirectly, either via an intermediate person or via contaminated fomites. Sukhrie *et al.* [140] reached the same conclusion, although neither their or our study included environmental sampling nor sampling from staff.

One of the clusters described in this chapter, Cluster 11 (GII.Pe_GII.4), was at the time suspected to be an outbreak by classical Infection Control investigations (14-09 B). This outbreak occurred on a bone marrow transplant ward; patients are in isolated rooms and have no direct contact (**Figure 5.27**). Indirect transmission was suspected, prompting a request for stool samples from parents on the ward. Of the three parents that submitted stool specimens, one was norovirus positive and part of Cluster 11 (NORO/51, father of the index patient NORO/52). It is not known whether the father contributed to the transmission chain; however it demonstrates that transmission does occur beyond the affected patients and provides a potential route for indirect transmission via visitors.

Figure 5.27 Layout of bone marrow transplant ward at GOSH in which norovirus outbreak occurred (Cluster 11). Each room is in isolation with its own bathroom facilities and positive pressure ventilated lobby. Blue arrows indicate direction of airflow, designed to prevent airflow from room to corridor and vice-versa.



Another potential source of indirect transmission between patients is infected staff; be it clinical or auxiliary. It is well documented that staff are affected during norovirus outbreaks; over a seven year period in five hospitals in England, on average the ratio of affected staff to affected patients was 1:1.75 (staff:patients), ranging from 1:7 in 2007/8 and 2008/9 to 1:1.15 in 2010/11 when almost as many staff were affected as patients [201]. Exclusion of affected staff can reduce transmission to patients [97] which is why national guidelines recommend symptomatic staff are excluded from work for 48 hours after resolution of symptoms [88].

An alternative route of indirect transmission is fomites contaminated with norovirus, which are a recognised source of infection that does not require direct contact between affected persons [202]. Contamination can occur in the environment immediately surrounding the patient, such as the bedside and in bathroom facilities, but also on patient notes trolleys, computer keyboards and portable medical equipment such as blood pressure equipment, pulse oximeters and tympanic thermometers [95]. During a norovirus outbreak in a long-term care facility, norovirus was also detected on an elevator button which was only accessed by staff [203]. Furthermore it is predicted that the transfer of norovirus is possible even after the virus has dried on hands or contaminated surfaces [204]. In the context of out-patients in a transmission chain fomites may become contaminated which are then touched by staff who return to the

wards, or the contaminated fomites themselves, such as patient notes or portable equipment, may be taken to the ward.

5.4.5 Utility of full genome sequencing and phylogeny for outbreak investigations

Full genome sequencing has proven in this study to be useful both by discriminating between infected patients based on genotype, for which both polymerase and capsid genotypes are critical, and by using phylogenetic analysis to link patients that were missed by classical Infection Control investigations.

Based on this, real-time sequencing of all new norovirus cases to support infection control investigations is useful and worthwhile. However currently the fastest achievable turn-around time for full genome sequences is one week and due to batching requirements often closer to 3 weeks (described in more detail in Chapter 4); whilst this is an excellent turn-around time for generating the associated volume of data, it is of limited use in the context of norovirus outbreaks which have incubation times as short as 12 hours. Therefore at this time infection control decisions must often be made in the absence of sequencing data; however newer and faster sequencing technologies are on the horizon which could provide data in a clinically useful time-frame. For instance the Oxford Nanopore Technologies MinION claims a sequencing speed of 500 bases per second; theoretically this could generate a norovirus full genome with 100-fold read depth in 25 minutes. The compatibility of platforms such as the MinION with SureSelect target enrichment is yet to be assessed.

To this end the data from this study should be used to improve and inform best practice when investigating norovirus transmission using classical epidemiology, for instance by considering norovirus positive patients on different wards during outbreak investigations, in particular if there are shared clinical teams. Sequencing data going forward should be used to confirm the findings of the infection control team and to identify gaps, as has been the case in this study, for the benefit of future investigations. In the case of outbreaks that last longer than a week or two, sequence data may be able to provide information in real-time.

5.4.6 Persistence vs. re-infection

Patients for longitudinal sampling were chosen at random, with a bias towards those who appeared to be chronically infected or who were norovirus PCR positive following a period during which norovirus could not be detected in stool, thus suggesting a new

infection. This was not a comprehensive collection of all such patients; consequently an accurate rate of re-infection vs. persistence cannot be determined. Nonetheless our results do suggest that patients are more likely to be re-infected with a new genotype if a period of negative stools occurs compared to those who were continuously PCR positive (67% and 17% re-infected, respectively), however re-infection can occur in as little as 58 days with no negative period in between (Px 73). This demonstrates super-infection with a different genotype is possible, as is seen in the last sequenced sample for Px 65, in whom a mixed infection with GII.3 and GII.4 is detected despite continuously PCR positive stools. As this was a retrospective study data is not available to determine whether super-infection is associated with worsening symptoms. However super-infection is not associated with a longer duration of infection therefore may not be clinically significant. As there were only three patients who had super-infections, a larger sample size is required to verify the clinical significance.

All, but one, patients who remained infected with the same genotype were shown by phylogenetic analysis to have persistence with the same virus, not re-infection with another strain of the same genotype. In the only patient to be re-infected with a different strain of the same genotype (Patient 63, genotype GII.P21_GII.3) over a year (466 days) had passed between the two norovirus episodes.

These results suggest that recent or current norovirus infection is not necessarily protective against re-infection with a different genotype. Norovirus animal and human vaccine studies to date have only included one or two genotypes, most commonly GI.1 and GII.4 [101], therefore this could have implications for vaccine development, especially considering the breadth of genotypes seen in our paediatric population. The population described in this chapter includes patients with immunosuppression and/or other co-morbidities, therefore the findings cannot be extrapolated to a general paediatric population; however it is these at-risk patients to whom vaccination programmes should be targeted, therefore the findings are of interest.

5.4.7 Conclusions

Classical epidemiology alone, i.e. linking patients in time and place, is not sensitive enough to identify all transmission events. The majority that are missed are either on a different ward, spread over time or the clusters are small (2 patients) therefore do not raise alarm bells. Conversely, molecular epidemiology is able to identify many transmission events that are missed using classical epidemiology, including small clusters of two patients with limited onward transmission and clusters with rumbling

transmission over several months. Moreover sequence analysis reveals the true extent of outbreaks which are only partially recognised by classical epidemiology, primarily due to episodes occurring on separate wards. In these instances transmission is likely to be indirect via staff, visitors or fomites.

Both capsid and polymerase genotyping are required as an adjunct to classical epidemiology; norovirus episodes linked in time and place and caused by the same genotype are likely to be linked. However phylogenetic analysis is necessary to confirm or exclude transmission between patients infected with the same genotype, in particular when separated in time or place. The polymerase and capsid shell domain sequences used for genotyping are not discriminatory enough for outbreak investigations [145]; in an immunocompetent population sequences of the P2 domain may suffice, however in an immunocompromised population full genome sequences are required. It is important however to take into account whether a plausible link between patients exists to confirm the occurrence of transmission; this must include attendance at outpatient clinics and other recent admissions.

5.5 ACKNOWLEDGEMENTS

As detailed in Chapter 4, I extracted RNA and prepared cDNA for 144 stool samples. For the remaining 210 samples, RNA from 116 samples (GOSH) was extracted by Divya Shah, under my supervision, and the remaining 94 samples from NNUH were extracted by NNUH Microbiology. cDNA for the remaining 210 samples was prepared by the UCL Pathogen Genomics Unit (PGU).

Sequencing library preparation and Illumina sequencing was done by the UCL PGU.

Full genome assembly of 201 samples was done by me. Assembly of the remaining 153 samples was done by the UCL PGU. All other analysis, including phylogenetic analysis, was done by me.

CHAPTER 6

ORAL RIBAVIRIN THERAPY FOR CHRONIC NOROVIRUS INFECTION IN IMMUNOCOMPROMISED PATIENTS

6.1 INTRODUCTION

Norovirus infections are self-limiting in immunocompetent hosts, with limited morbidity aside from dehydration. In immunocompromised patients however, there is a risk of chronic infection with significant associated morbidity. Chronic infections are bi-phasic [177] with an acute phase of vomiting and diarrhoea, followed by chronic viral shedding and diarrhoea lasting weeks to years. The stool volume can amount to several litres a day, which is comparable to that observed with cholera infections [205]. The majority of case reports describe patients to be symptomatic during this extended period of shedding, with up to 24 bowel movements per day [206], however chronic norovirus infections can experience intermittent symptoms of diarrhoea [116] or be asymptomatic [207].

Chronic infection can lead to dehydration, malnutrition, dysfunction of intestinal barrier [80, 207-209], dramatic weight loss [54, 81, 82, 118, 120, 207, 210-217] and a requirement for nutritional support [54, 120, 208, 209, 211, 212, 215-217]. The requirement for nutritional support may be proportional to the duration of infection, as suggested by a cohort of hematopoietic stem cell transplant (HSCT) recipients in whom patients that excreted virus for more than 100 days required nutritional support for 165 days on average compared to those who excreted virus for less than 100 days required intravenous nutritional support for a shorter period, on average 46 days [54]. Prolonged norovirus infections can lead to worse outcome of underlying disease [80] or graft rejection in solid organ transplant recipients [218]. In extreme cases, norovirus in immunocompromised patients can be fatal; one third of fatal norovirus cases are reported to occur in organ transplant recipients or in patients receiving chemotherapy for malignancy [219].

The prevalence of norovirus infections in immunocompromised patients is not formally established, however a small number of case reports suggest it is around 20%. In a cohort of 61 haematopoietic stem cell transplant (HSCT) recipients, 21% developed norovirus infections [54]. In a cohort of 78 renal transplant recipients, 17% were infected with norovirus [116]. In cohorts with mixed causes of immunodeficiency, 22% of 116 patients with a history of bone marrow transplant (BMT) or HSCT developed norovirus infection [118] and 23% of 47 patients who had undergone BMT, solid organ transplant or who were immunosuppressed due to an oncological diagnosis [117] had norovirus infections.

Severe, protracted diarrhoea caused by norovirus has been reported in several immunocompromised populations, including HSCT recipients [54, 80, 120, 176] who

remain immunosuppressed until immune reconstitution post-transplant. Solid organ transplant recipients [218], namely intestinal transplant [220], pancreatic transplant [221], combined liver, pancreas and bowel transplant [209, 222], kidney transplant [81, 116, 207, 210, 223], lung transplant [224, 225] and heart transplant [211-213], remain on life-long maintenance immunosuppression to prevent graft rejection. Consequently solid organ recipients are at risk of chronic norovirus infections at any time post-transplant, with reported infections occurring anywhere from one month to 10 years post-transplant. In HSCT recipients on the other hand, immunosuppression is largely temporary therefore the majority of reported infections occur within the first year post-transplant. Severe norovirus infections are also described in patients with malignancy [82, 206, 208, 226-228] who are immunosuppressed either as a direct consequence of the malignancy or as a consequence of therapy, and in acquired immune deficiency syndrome (AIDS) due to HIV infection [214]. Patients with primary immunodeficiencies are also described as at risk of chronic and debilitating norovirus infections [84, 195, 215-217, 229]; these patients are permanently immunosuppressed due to genetic abnormalities. These case reports are summarised in **Table 6.1**.

Viral clearance is reportedly associated with immune reconstitution and T cell recovery in animal models [55] and human case reports [54, 228]. Reducing or withholding immunosuppression has been shown to improve clinical symptoms caused by norovirus in 2 days to 5 weeks [116, 120, 207, 220, 230], however in a renal transplant cohort the intensity of immunosuppression correlated with diarrhoeal symptoms but not with viral shedding [116]. In an HIV infected patient symptoms caused by norovirus infection improved once CD4+ count increased [214]. However, reduction of immunosuppression to promote norovirus clearance is not always possible; in solid organ transplant recipients it risks graft rejection, in bone marrow transplant recipients one has to wait for immune reconstitution and in primary immunodeficiencies immunosuppression is not an induced state therefore cannot be changed. In these scenarios specific antiviral therapies to clear chronic norovirus infections and aid recovery of gut function would be desirable and possibly reduce chronic norovirus-associated morbidity.

Table 6.1 Summary of chronic norovirus cases in immunosuppressed patients in the literature

Reason for immunosuppression	Reference	N	Adult or paed.	Time since transplant	Duration of shedding	Bowel movements per day	Associated morbidity (additional to diarrhoea)	Treatment
Malignancy	Ghosh *et al.* (2016) [206]	8	A and P	N/A	46–270 days	1–24	Not stated	Treated 17 norovirus cases (including non-chronic) with nitazoxanide. Mean duration of diarrhoea in treated (n=17) 7 days (3–30), untreated (n=32) 12 (2–32) (P=0.09)
	Knoll *et al.* (2016) [228]	1	A	N/A	Not stated	10–15	Not stated	Nitazoxanide for 7 days ineffective. IVIG ineffective. Recovery associated with immune reconstitution
	Bruggink *et al.* (2015) [226]	1	A	N/A	>241 days	Not stated	*No sequence changes in ORF1 and ORF 2 over 241 days*	None
	Ronchetti *et al.*(2014) [227]	1	A	N/A	12 weeks	4–7	Not stated	Parenteral and oral immunoglobulin (one day of each) unsuccessful
	Ludwig *et al.*(2008) [82]	9	P	N/A	22–433 days (median 46)	Not stated	Two patients with excretion >100 days had weight loss, one with severe growth retardation	None
	Simon *et al.* (2006) [208]	5	P	N/A	37–140 days	Not stated	35% required parenteral nutrition (7/20 infected patients) 1 patient: life-threatening bleeding from lower gastrointestinal tract 1 patient: failure to thrive, dehydration, transient lactose malabsorption. Several months of protracted diarrhoea requiring parenteral nutrition. Mucosal biopsies showed flattened mucosa with patchy haemorrhagic areas	None

Reason for immunosuppression	Reference	N	Adult or paed.	Time since transplant	Duration of shedding	Bowel movements per day	Associated morbidity (additional to diarrhoea)	Treatment
Lung transplant	Gairard-Dory *et al.* (2014) [224]	12	A	25 months (±18)	Not stated	Mean 7 ±3	52% had diarrhoea-induced renal impairment	All treated with oral immunoglobulin for 2 days; 11/12 success, 1/12 mildly improved. NB success = resolution of clinical symptoms, no follow-up norovirus testing
	Boillat Blanco *et al.* (2011) [225]	1	A	2 weeks post-lung, 5 years post-HSCT	6 months	Not stated	Not stated	Resolved with switch from sirolimus to tacrolimus
Pancreatic transplant	Echenique *et al.* (2016) [221]	1	A	841 days	2543 days	20	Hospitalised for intravenous rehydration	None
Renal transplant	Coste *et al.* (2013) [210]	15	A	9.5 years (±7.7)	Not stated	Mean 6 (±3)	Greater weight loss compared to other causes of diarrhoea	None
	Roos-Weil *et al.* (2011) [81]	15	A	37 months (±37)	Median 289 (107–581)	Not stated	Greater weight loss compared to other causes of diarrhoea 9-fold longer duration of symptoms compared to other causes of diarrhoea Diarrhoea-associated acute renal failure in 81%	None
	Schorn *et al.* (2010) [116]	9	A	Median 42 months (1–136)	97–898 days	Not stated	5/9 had intermittent symptoms 5/9 had to be hospitalised Decrease in renal function in all (reversible)	Treated with cautious reduction in immunosuppression. Intensity of immunosuppression correlated with diarrhoeal symptoms but not with viral shedding (all patients improved clinically but only 3/9 stopped shedding)

Reason for immunosuppression	Reference	N	Adult or paed.	Time since transplant	Duration of shedding	Bowel movements per day	Associated morbidity (additional to diarrhoea)	Treatment
	Chagla *et al.* (2013) [223]	1	A	5 years post-pancreatic, 15 years post-renal	18 months (79 days confirmed by PCR)	5–10	Hospital admission for diarrhoea Diarrhoea required placement of a rectal tube	Parenteral immunoglobulin unsuccessful Enteral Immunoglobulin for 2 days successful
	Westhoff *et al.* (2009) [207]	2	A	3 months and 14 years	5.5 months and 3 months	Not stated	Both admitted to hospital due to diarrhoea 1 patient: Intermittent diarrhoea for 3 months, >5 kg weight loss, shortened and broadened villi on biopsy	1 patient: spontaneous resolution 1 patient: symptom resolution and cleared virus once withheld tacrolimus
Heart transplant	Ebdrup *et al.* (2011) [211]	1	A	3 years	56 days	10–15	Administered intravenous fluids Parenteral nutrition for 12 days Severe weight loss (9 kg) Required renal dialysis (severe metabolic acidosis)	Immunoglobulin via duodenal tube successful
	Nilsson *et al.* (2003) [212]	1	A	6 years	>3 years	4–8	Admitted to hospital for intravenous rehydration Severe weight loss Parenteral nutrition 5 days per week for > 3 years *Longitudinal sampling showed one permanent amino acid change and one transient amino acid change in polymerase after 1 and 4 months, respectively.*	Breast milk (three times per day) for 3 weeks, oral immunoglobulin for 15 days, IV immunoglobulin for 5 days and reduction of immunosuppression; all unsuccessful.

Reason for immunosuppression	Reference	N	Adult or paed.	Time since transplant	Duration of shedding	Bowel movements per day	Associated morbidity (additional to diarrhoea)	Treatment
	Engelen *et al.* (2011) [213]	1	A	14 months	365	Not stated	10 kg weight loss	Symptom and virus resolution when everolimus added to immunosuppressive regimen
Intestinal transplant	Kaufman *et al.* (2003) [220]	1	P	178 days	>120 days	Not stated	Intravenous fluid therapy required	Recovery associated with reduction in immunosuppression
Combined liver, pancreas and small bowel transplant	Lee *et al.* (2008) [209]	1	P	61 days	114 days	Not stated (high ileostomy losses)	Haemorrhagic necrosis of small bowel Total parenteral nutrition required	
	Florescu *et al.* (2008) [222]	2	P	4 months and 7 months	2–8 months	Not stated (increased ileostomy output)	Aggressive fluid replacement required	Oral human serum immunoglobulin for 2 days successful
Mixed solid organ transplants (kidney, liver, pancreas, small bowel or heart)	Lee *et al.* (2016) [218]	67	A	Median 2.82 years (12–5945 days)	Median 135 days (1–2569 days)	Not stated	Duration of diarrhoea significantly longer compared to non-norovirus causes (median 135 vs. 25 days), 88% developed chronic diarrhoea 39/67 patients hospitalised for diarrhoea 1 diarrhoea-related death, 2 diarrhoea-related graft failures 2 kg weight loss at 1 month	None
Mixed solid organ transplants	Avery *et al.* (2017) [231]	31	A	37 months (<1–312)	4 months (<1–20)	Not stated	Wasting (>10% of body weight) (35%) Acute kidney injury (23%)	All resolved, unclear which therapy responsible (16% spontaneous; 84% on nitazoxanide, reduced immunosuppression and/or IVIG)

Reason for immunosuppression	Reference	N	Adult or paed.	Time since transplant	Duration of shedding	Bowel movements per day	Associated morbidity (additional to diarrhoea)	Treatment
Hematopoietic stem cell transplant (HSCT)	Saif *et al.* (2011) [54]	13	P	−80–63 days	Median 150 days (60–380)	Not stated	All required nutritional support Patients excreting >100 days: required combined nutritional support for 165 days (35–416) Patients excreting <100 days required nutritional support for 46 days (26–71) Weight loss	Viral clearance correlated with T cell recovery
	Roddie *et al.* (2009) [120]	6	A	Median 10.5 months (0.25–60 months)	Median 7 months (3–14 months)	Not stated	All chronically infected (6/12) required enteral or parenteral nutrition 1 patient 14 kg weight loss (25% of body weight) 1 patient died of malnutrition 26–108 days hospitalisation (median 73)	2 patients responded to withdrawal of immunosuppression
Mixed case series: solid organ and HSCT (not specified)	Ye *et al.* (2015) [118]	11	P	Median 137 days (−3–1962)	Median 12.5 (1–324)	Median 9 (3–16)	55% hospitalised for diarrhoea 27% required ICU admission (compared to 0% non-norovirus; P=0.02) Median 1.6 kg weight loss (vs. 0.6kg non-noro P<0.01)	
HIV positive	Wingfield *et al.* (2010) [214]	1	A	N/A	7 months	20	Weight loss 3 stone	3 days IVIG unsuccessful Clearance associated with increase in CD4 count

Reason for immunosuppression	Reference	N	Adult or paed.	Time since transplant	Duration of shedding	Bowel movements per day	Associated morbidity (additional to diarrhoea)	Treatment
Primary immunodeficiency - CVID	Woodward *et al*. (2015) [215]	8	A	N/A	Up to 22 months	Up to 10–20	Weight loss, 5/8 required long-term parenteral nutrition	5/8 treated with ribavirin; 2/5 successfully, 3/5 unsuccessfully
	Duraisingham *et al.* (2015) [216]	3	A	N/A	9–30 months	3–4 (only stated for 1 patient)	All severe weight loss, 1 patient malnutrition and long term parenteral nutrition	1 patient nitazoxanide (10 days) unsuccessful 1 patient 2 weeks ribavirin unsuccessful 2 patients oral immunoglobulin, 1 with symptom improvement but not viral clearance
	Van de Ven *et al.* (2011) [217]	1	P	N/A	>2 years	Not stated	Severe enteropathy with 19% weight loss, required parenteral nutrition	Unsuccessful IVIG, duodenal immunoglobulin, duodenal breast milk, ribavirin (10 days),
Primary immunodeficiency - other	Frange *et al.* (2012) [84]	11	P	N/A	Up to 10 months	Not stated	Mild to severe dehydration and denutrition, exacerbation of immune-mediated enteropathy	IVIG (3/11 patients) and oral immunoglobulin (1 patient) unsuccessful
	Gallimore *et al.* (2004) [229]	1	P	N/A	156 days	Not stated	None reported	None
	Xerry *et al*. (2010) [195]	2	P	N/A	48–69 days	Not stated	Not stated	None

N, number of patients; A, adults; P, paediatric; N/A, not applicable; ±, standard deviation

CVID, common variable immunodeficiency; Other, SCID (severe combined immunodeficiency), cartilage hair dysplasia, Wiskott-Aldrich syndrome, MHC Class II deficiency, agammaglobulinaemia, chronic granulomatous disease

The development of specific antivirals against norovirus infections has been somewhat neglected in comparison to other viral pathogens, primarily due to the acute short-lived nature of infection in immunocompetent hosts, and thus the majority of people. However norovirus is increasingly recognised as a major cause of morbidity, and in rare cases mortality, in immunocompromised patients, stimulating research activity to develop an effective antiviral therapy; nonetheless these studies remain few and on a small scale. Several compounds have been shown to have some degree of antiviral activity against murine norovirus or human norovirus replicon systems [232]. Of these there are two that are already licenced for the treatment of other infections in humans and therefore available to use clinically. These are nitazoxanide, licenced for *Cryptosporidium* and *Giardia* infections, and ribavirin, licenced for the treatment of respiratory syncytial virus (RSV) in infants and children and Hepatitis C virus (HCV) infections. In HCV infections ribavirin is given by mouth with peg interferon alpha or interferon alpha. In RSV bronchiolitis ribavirin is administered by inhalation, although there is no evidence that ribavirin produces clinically relevant benefit in RSV bronchiolitis [233].

Ribavirin is a purine mimic synthetic nucleoside analogue that resembles guanosine [234]. First synthesised in 1972 [235, 236] it is a broad range antiviral with activity against RNA and DNA viruses; *in vitro* activity has been demonstrated against 36 virus species across 18 genera and ten families [237].

Ribavirin has been shown to have *in vitro* dose dependent antiviral activity against murine norovirus and norovirus replicon-bearing cells [238]. More recently successful treatment with ribavirin has been reported in two adult Common Variable Immunodeficiency (CVID) patients with chronic norovirus infection, in whom 6 and 14 months of ribavirin therapy resulted in clearance of norovirus, complete resolution of clinical symptoms, improved nutritional status and almost complete recovery of duodenal villous morphology. Conversely, three patients from the same CVID cohort were treated with ribavirin for 9, 20 and 21 months without clearance of norovirus [215]. A possible side-effect or ribavirin is reversible haemolytic anaemia caused by the accumulation of ribavirin in erythrocytes.

There are several postulated mechanisms of action of ribavirin [239], which are summarised in **Table 6.2**. One which is measurable *in vivo* is the accumulation of mutations leading to lethal mutagenesis. RNA viruses lack proof-reading enzymes, resulting in a high mutation frequency of 10^{-3} to 10^{-5} substitution per nucleotide per replication cycle [21, 51]. Intra-host viral populations consequently exist as a heterogeneous population with a consensus sequence and minority variant sequences,

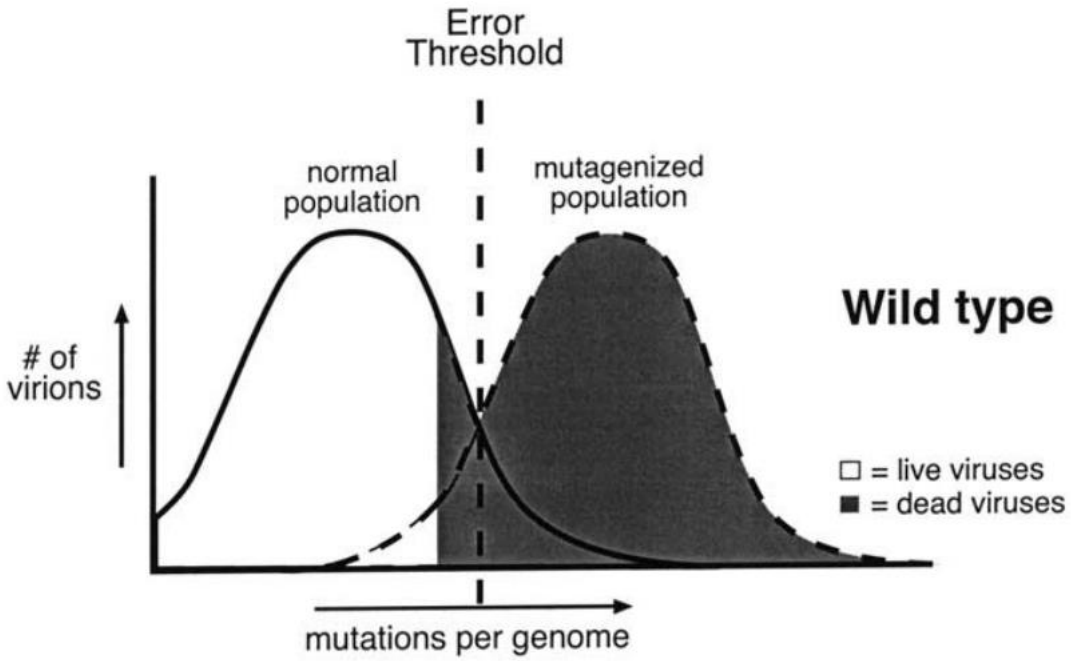
known as quasispecies. This high mutation frequency is advantageous as it allows rapid evolution and adaptation, however there is an upper limit of genome variability, known as the error threshold, beyond which errors become deleterious to the population. RNA viruses exist at the edge of the error threshold [240-242]; an increase in mutation frequency forces the population beyond the error threshold [242] thus is detrimental (**Figure 6.1**).

Table 6.2 Mechanisms of action of ribavirin against RNA viruses, as reviewed by Graci & Cameron [239]

Mechanism of action	Outcome
REDUCTION OF CELLULAR GTP POOLS RMP binds to the active site of IMPDH, the cellular enzyme involved in the first step of guanosine nucleotide synthesis, thus competitively inhibiting cellular GTP synthesis.	Inhibits translation, transcription and replication
IMMUNOMODULATORY EFFECT Mechanisms not understood	Enhances T-cell response
INHIBITION OF RNA CAPPING (7-METHYL GUANOSINE) Competitive inhibition of capping enzymes *or* Incorporated as a cap analogue	RNA not competent for translation
INHIBITION OF RNA POLYMERASE Competition with GTP and ATP pools so RTP incorporated into RNA *or* Binds to RdRp close to active site and changes conformation	Reduced elongation activity if RBV in template
MUTAGENESIS RTP has equal binding efficiency with cytidine or uridine Potential for mutagenesis due to incorporation of RTP during replication, compounded by ambiguous base pairing with CTP or UTP during next round of replication	Accumulation of mutations leading to lethal mutagenesis

GTP, guanosine triphosphate; RMP, ribavirin monophosphate; IMPDH, inosine monophosphate dehydrogenase; RdRp RNA dependent RNA polymerase; ATP; adenosine triphosphate; RTP, ribavirin triphosphate; CTP cytosine triphosphate; UTP, uracil triphosphate

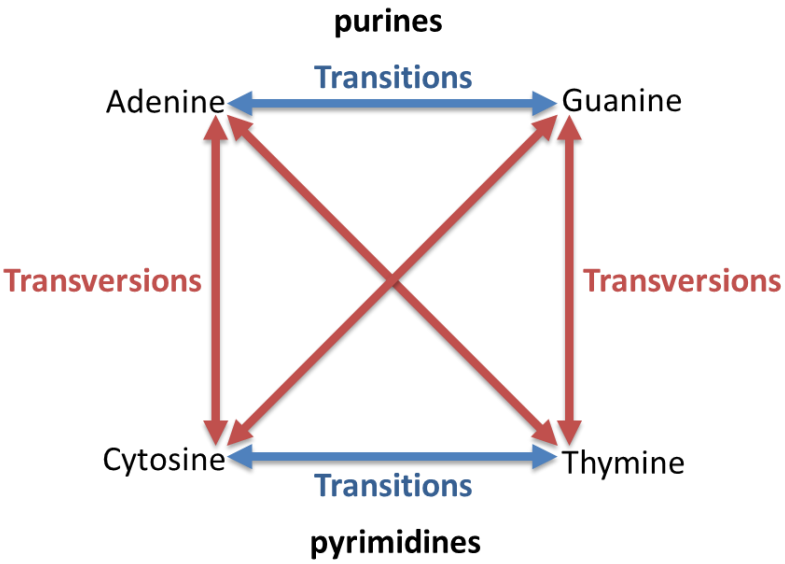
Figure 6.1 Model of viral quasispecies at the error threshold of lethal mutagenesis. The majority of viruses, having fewer mutations, are viable. Some viruses, with a greater number of mutations, are non-viable. A small increase in mutation frequency is predicted to push the virus population into lethal mutagenesis (mutagenized population), where the number of errors per viral genome is sufficiently high to lethally mutate the majority of the virus populations. Figure reproduced, with permission, from Vignuzzi *et al.* [237]



Lethal mutagenesis has been demonstrated to be the primary mechanism of action of ribavirin against Hantaan virus, Hepatitis C virus, foot and mouth disease virus and poliovirus [237]. In poliovirus, an increase in mutation frequency beyond the error threshold results in reduced virus infectivity [237], demonstrating that the accumulation of deleterious mutations due to lethal mutagenesis caused by ribavirin ultimately leads to loss of fitness. To date, virus mutation frequencies have been determined by PCR amplification of a region of the genome followed by cloning and capillary sequencing of 70 or more colonies [158].

Since ribavirin is a purine mimic which binds with equal efficiency to cytidine or uridine, ribavirin driven mutations are expected to be transitions rather than transversions (**Figure 6.2**).

Figure 6.2 Schematic of transition and transversion nucleotide substitutions



This chapter reports the use of ribavirin treatment in two child recipients of organ transplants in whom chronic norovirus infection was associated with malnutrition and weight loss. In order to determine whether ribavirin was acting on norovirus, deep sequencing was used to look for signatures of ribavirin mutagenesis in virus sampled pre– and post–treatment in relation to the clinical outcome following treatment.

6.2 MATERIALS AND METHODS

6.2.1 Case history and ribavirin therapy

Patient A was a six year old girl who had a heart transplant at one month of age, following enteroviral myocarditis (coxsackie virus). She remained on long-term immunosuppressive therapy with sirolimus following transplant. In August 2012 she developed severe diarrhoea, which was diagnosed by routine real-time RT-PCR as norovirus GII in September 2012. By November 2012 she was malnourished, had lost a third of her body weight (from 17.5 kg in July 2012 to 12.8 kg in November) and was unable to attend school due to the severity of diarrhoea.

Three months after norovirus was first detected, oral ribavirin therapy for Patient A was commenced at 2.5 mg/kg BD for 4 weeks, with no evidence of adverse effects. The dose was increased to 5 mg/kg BD for a further 4 weeks. Therapy was stopped after a total of 2 months. Four weeks prior to starting Ribavirin therapy, Patient A's sirolimus dose (immunosuppressant used to prevent organ rejection) was decreased from 1 to 0.5 mg OD and further reduced to 0.2 mg OD ten days prior to starting ribavirin therapy.

Patient B was a three year old boy with congenital nephrotic syndrome who received a kidney transplant at 22 months of age, following which he remained on long term immunosuppressive therapy with tacrolimus. He developed chronic diarrhoea a few days after transplant which was diagnosed at the time as norovirus GII.

Oral ribavirin therapy for Patient B was commenced 7 months after norovirus was first detected at 2.5 mg/kg BD for 5 weeks; the dose was then increased to 5 mg/kg BD and to 7.5 mg/kg seven weeks after that. Ribavirin was well tolerated and he remained on treatment for a total of 5 months. Patient B's tacrolimus (immunosuppressant) dose was not changed throughout the study period.

Stool frequency and consistency was recorded daily by parents commencing from before and continuing for a few weeks after ribavirin treatment and weight was measured weekly. Stool samples were collected once to twice weekly to monitor norovirus detection; blood samples were collected once weekly to monitor for anaemia associated with ribavirin therapy and every two weeks to monitor plasma ribavirin levels.

6.2.2 Viral load monitoring

Stool samples from each patient were collected and tested for norovirus once to twice per week using real-time PCR, the methods for which are described in Chapter 2. Copy number was determined using a standard curve, as described in Chapter 2. To generate a viral load in copies/gram, the amount of stool used to make the stool lysate was weighed.

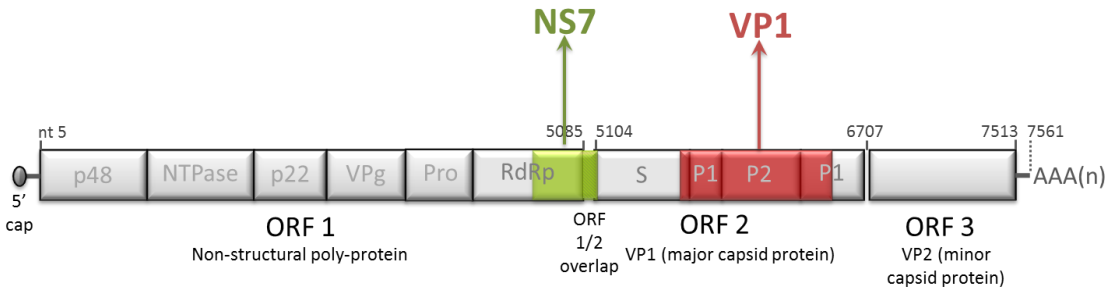
6.2.3 cDNA synthesis

cDNA was prepared using universal norovirus GII.4 primers, as described in detail in Chapter 2.

6.2.4 Method to determine mutation frequencies: cloning vs. deep sequencing

Fragments of the RNA polymerase (NS7, 782 nucleotides) and capsid (VP1, 694 nucleotides) genes (**Figure 6.3**) were amplified from cDNA by PCR from samples collected from Patient A before and during ribavirin therapy. In parallel, VP1 and NS7 amplicons were cloned, with 96 colonies capillary sequenced, and deep sequenced on an Illumina MiSeq. Methods for cloning and MiSeq library preparation are described in Chapter 2.

Figure 6.3 Schematic of norovirus genome with NS7 (green) and VP1 (red) PCR target regions highlighted. Nucleotide numbering is based on a GII.4 genome



A norovirus GII.4 full length clone, constructed by the Goodfellow lab, was used as a control throughout the process.

6.2.5 Norovirus mutation frequencies pre- and post-ribavirin therapy

To estimate the mutation frequency of the total norovirus population in each sample, NS7 and VP1 amplicons were generated from longitudinal samples collected before and during ribavirin therapy from Patient A, Patient B and from a third and fourth child, Patient C and Patient D who were chronically infected bone marrow transplant recipients who were not treated with ribavirin. All amplicons were sequenced by capillary sequencing to determine the consensus sequence and deep sequenced on an Illumina MiSeq to determine the mutation frequency, as described in Chapter 2. A GII.4 full length clone, constructed by the Goodfellow lab, was used as a control throughout the process.

6.2.6 Calculation of mutation frequency

The mutation frequency is expressed as number of mutations per 10,000 nucleotides. Details of how this was calculated are described in Chapter 2.

6.2.7 Statistical analysis

All statistical analysis was undertaken in SPSS v23, as described in Chapter 2.

6.2.8 Full genome sequencing

Pre- and post-ribavirin samples for Patient A and Patient B were sequenced using SureSelect target enrichment to generate full genome consensus sequences (methods described in Chapter 2). Genomes were assembled using the de-novo pipeline described in Chapter 2. Full genome sequences were used to determine the full polymerase and capsid protruding domain sequences, as the NS7 and VP1 amplicons only provide partial sequences for these regions.

6.2.9 Mapping amino acid changes

Known epitopes were identified in the capsid protruding (P) domain amino acid sequences from Patient A and Patient B ten and 27 days into ribavirin therapy, respectively, as described in Chapter 2. Residue changes seen during ribavirin therapy were highlighted.

6.2.10 Phylogenetic analysis of capsid nucleotide sequences

A maximum likelihood tree was reconstructed for the VP1 PCR capsid sequences, as described in Chapter 2.

6.3 RESULTS

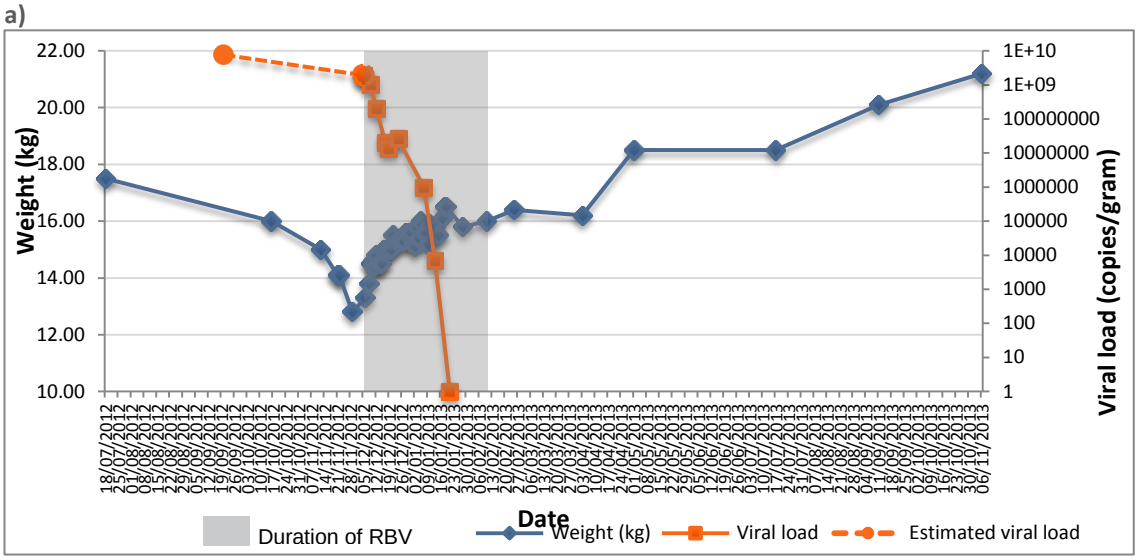
6.3.1 Viral load, weight and diarrhoea

During ribavirin therapy the stool viral load for Patient A decreased and became undetectable 6 weeks after start of therapy. During this time Patient A's weight steadily increased from 13.3 kg at the start of ribavirin to 16 kg two months after ribavirin was stopped. The patient was followed up for a further nine months during which time her weight continued to increase to 21.2 kg (11 months after start of therapy) and norovirus remained undetectable (**Figure 6.4a**).

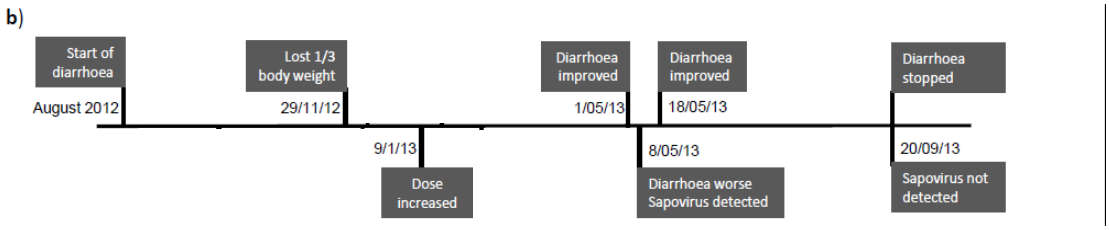
Patient A's stool output remained unchanged during ribavirin therapy with 6-10 liquid stools per day. Diarrhoea improved significantly two months after norovirus was eradicated and stopped completely after seven months (**Figure 6.4b**).

Figure 6.4 Patient A **a)** norovirus viral load and weight; **b)** clinical timeline

a)



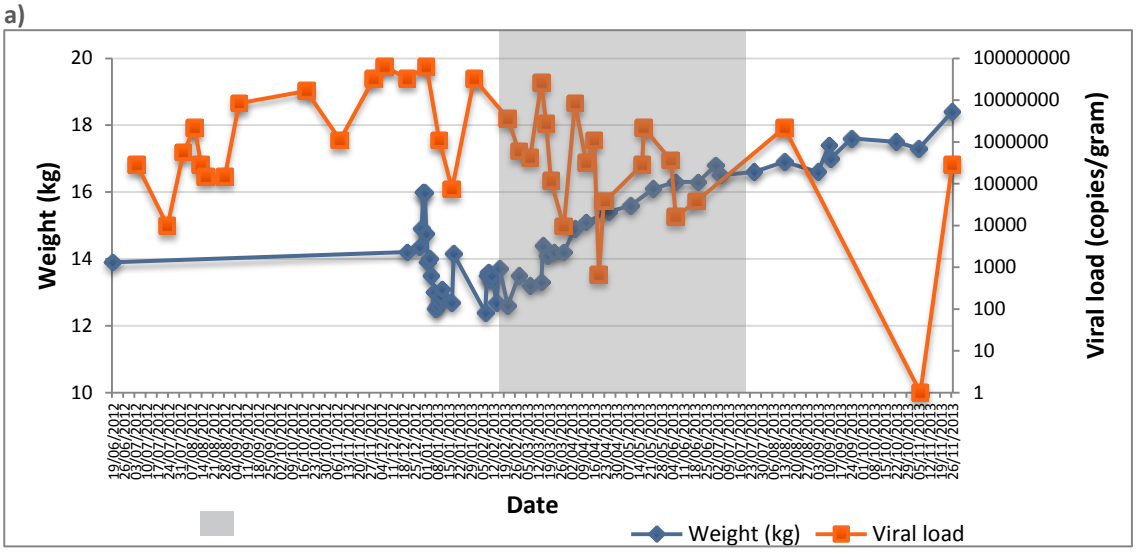
b)



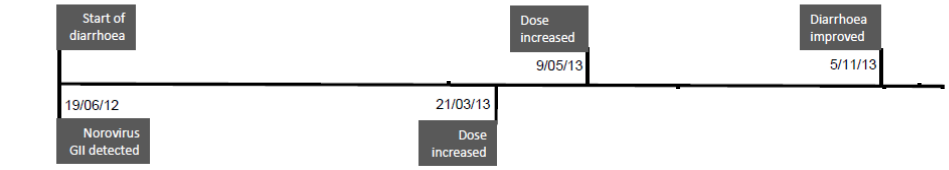
Prior to norovirus infection Patient B's weight was in the 91st centile. During norovirus infection Patient B suffered significant weight loss to 12.7 kg, on the 25th centile. During 5 months of ribavirin therapy Patient B also gained weight, from 12.7 kg to 16.5 kg, returning to the 91st centile. Patient B was followed up for a further three months during which time he continued to gain weight to 18.4 kg (**Figure 6.5**).

Figure 6.5 Patient B **a)** norovirus viral load and weight; **b)** clinical timeline

a)



b)



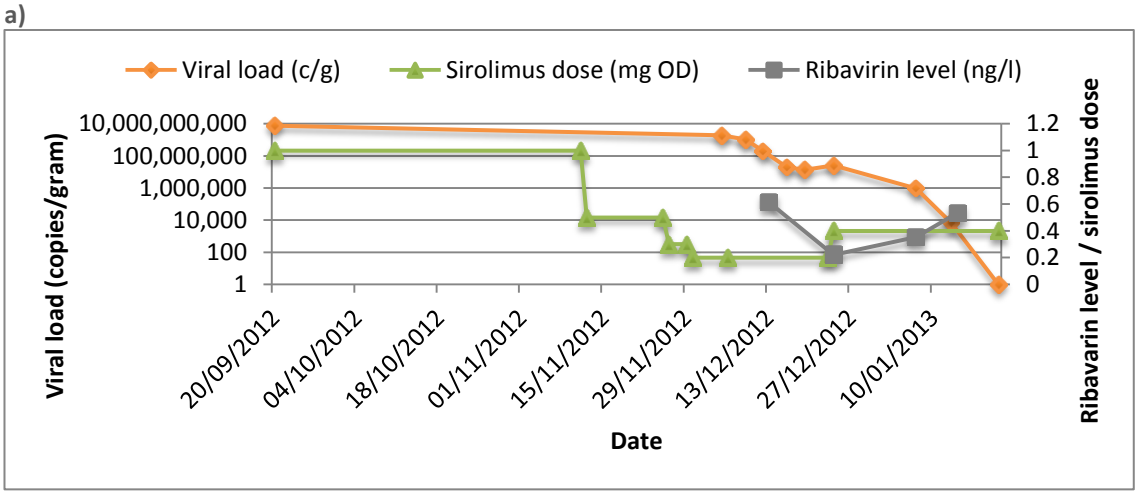
Patient B's viral load fluctuated during ribavirin therapy, but his stool output remained unchanged, with 1–9 liquid stools per day. Two months after the end of therapy diarrhoea had improved significantly and norovirus was no longer detectable. However 2 months later norovirus was once again detectable and two weeks later severe diarrhoea, with up to 13 stools a day, re-started, coinciding with an acute astrovirus infection. Once the astrovirus infection resolved spontaneously, norovirus remained detectable for a further 6 months however diarrhoea did not return.

6.3.2 Ribavirin levels

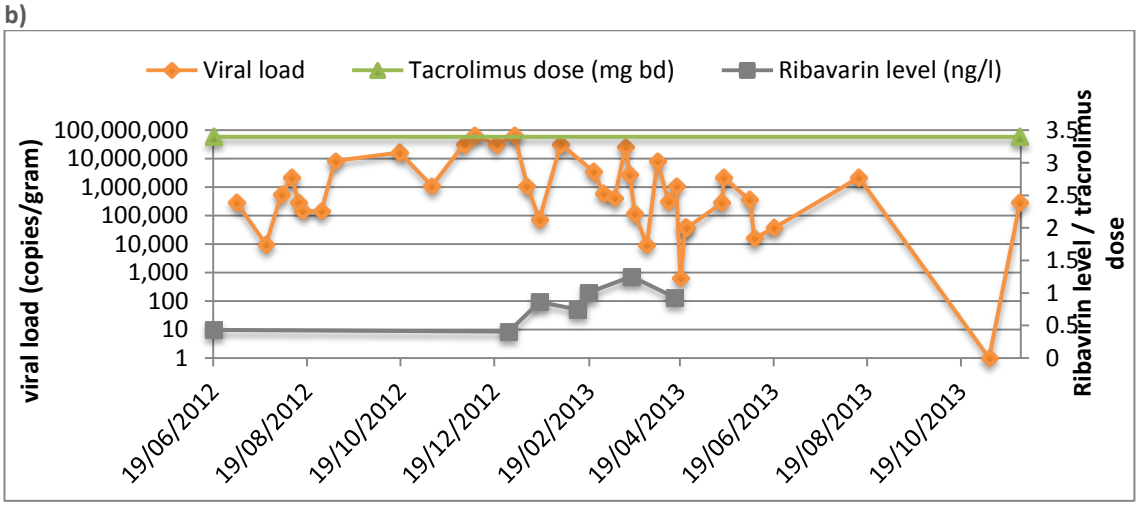
Ribavirin levels in plasma for Patient A peaked at 782 ng/ml and 1,252 ng/ml for Patient B (**Figure 6.6**).

Figure 6.6 a) Norovirus viral load, plasma ribavirin level and sirolimus dosage for Patient A; **b)** Norovirus viral load and plasma ribavirin level for Patient B (dosage of immunosuppressant unchanged)

a)



b)



Red blood cell counts for Patients A and B stayed within normal limits therefore there was no indication of anaemia during therapy, suggesting that ribavirin accumulation had no adverse effects.

6.3.3 Consensus sequences

6.3.3.1 Whole genome sequences and genotypes

Whole genomes sequences, used to derive the full RdRp and capsid protruding domain sequences, were successfully determined for Patient A and Patient B (**Table 6.3**).

Table 6.3 Assembly details of norovirus full genome sequences used to derive capsid P domain and polymerase sequences

Sample name	Assembly method	Assembled genome size	Estimated genome coverage	Average read depth
Patient A pre-ribavirin	De novo	7408	97.99%	20,548
Patient A 10 days post-ribavirin	De novo	7495	99.14%	10,388
Patient B pre-ribavirin	De novo	7466	98.76%	491
Patient B 27 days post-ribavirin	De novo	7499	99.19%	153

Genotyping using the Norovirus genotyping tool [4] and full genome sequences showed all norovirus positive samples from Patient A and Patient B to be GII.P4_GII.4 New Orleans_2009. The positive control used throughout this study was a GII.P4_GII.4 Farmington Hills_2002.

Genotyping using the Norovirus genotyping tool [4] and NS7 and VP1 amplicon sequences showed all samples from Patient C were GII.Pe_GII.4 Sydney_2012 and from Patient D were GII.P4 New Orleans_2009_GII.4 Sydney_2012.

6.3.3.2 Consensus sequence changes in RdRp

The second and subsequent samples collected from Patient A, seven days after starting ribavirin treatment, showed five coding changes in the RNA polymerase gene (**Figure 6.7**, **Table 6.4**). Although most substitutions did not alter the predicted protein properties, one (M231K) corresponds to a change from a very hydrophobic to a hydrophilic amino acid near the RdRp palm domain.

Figure 6.7 Coding changes in RNA polymerase gene of Patient A mapped onto RNA polymerase protein structure of Norwalk virus RdRp. Blue indicates residues known to be involved in catalysis. Red indicates changes selected with ribavirin treatment (figure prepared by Lucy Thorne, University of Cambridge)

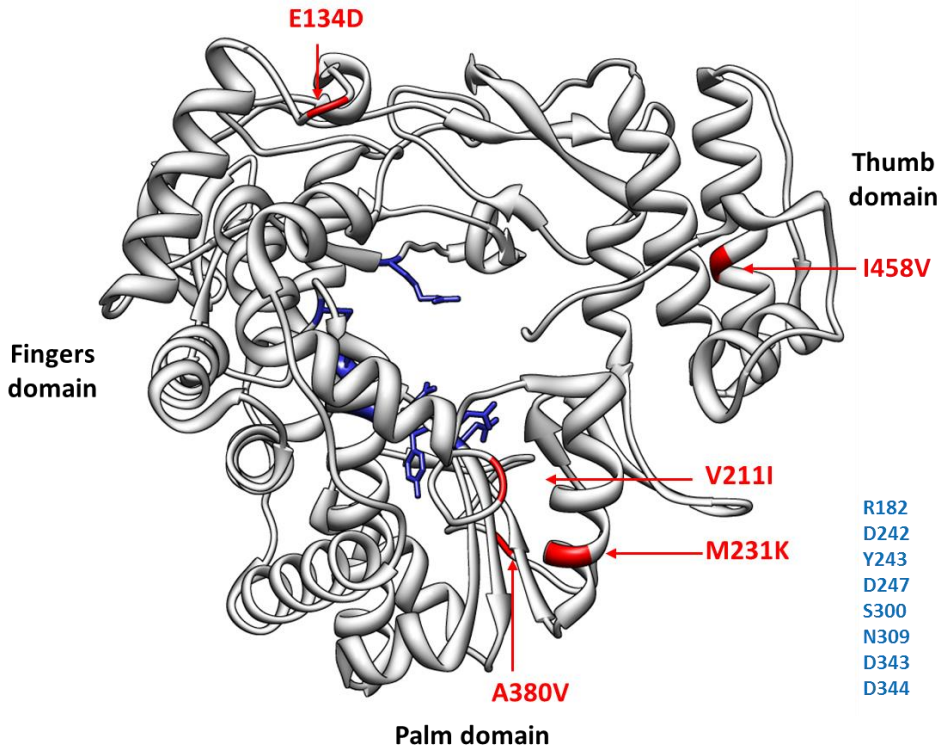
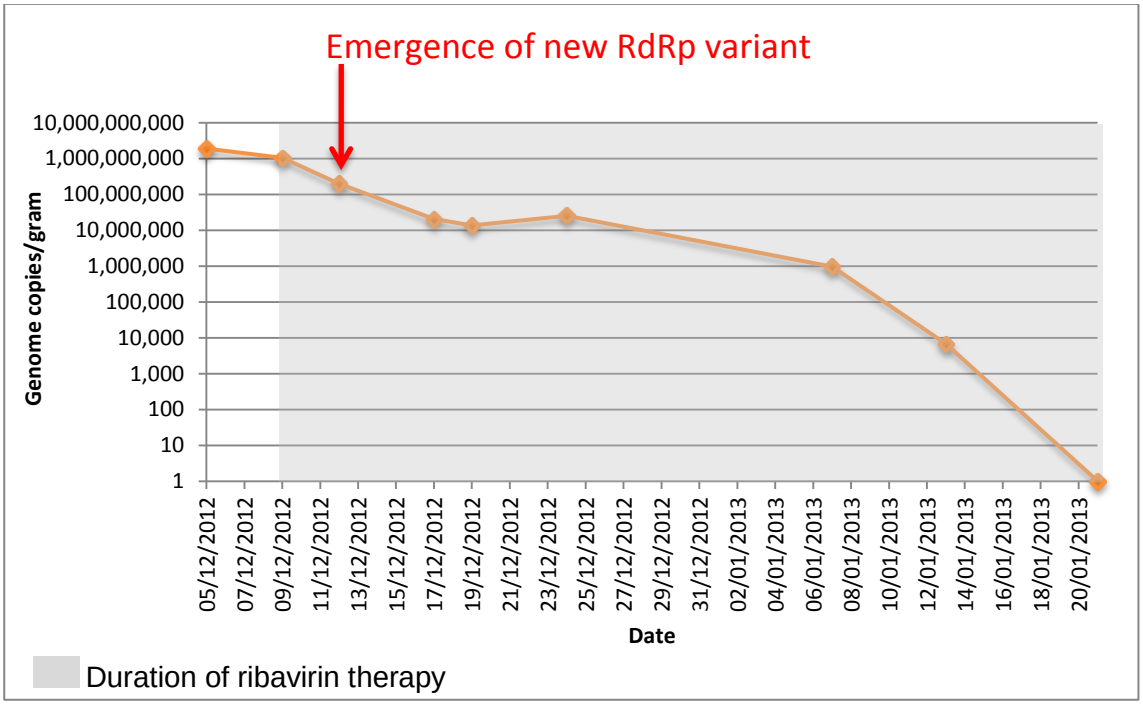


Table 6.4 Coding (non-synonymous) nucleotide changes in norovirus RNA polymerase gene from Patient A

	Nucleotide position				
Sample collection date	3973	4202	4263	4710	4943
29/11/12 (pre ribavirin)	G	G	T	C	A
05/12/12 (pre ribavirin)	G	G	T	C	A
16/12/12 (during ribavirin)	T	A	A	T	G
07/01/12 (during ribavirin)	T	A	A	T	G
13/01/12 (during ribavirin)	T	A	A	T	G
Amino acid change	E134D	V211I	M231K	A380V	I458V

The emergence of this RdRp variant coincided with a log reduction in norovirus viral load and subsequent clearance of the virus (**Figure 6.8**). Neither Patient B nor patient C or D showed consensus level coding changes in the RdRp gene.

Figure 6.8 Norovirus viral load in Patient A, with emergence of RdRp variant



6.3.3.3 Consensus sequence changes in capsid

Patients A and B had 10 and 11 amino acid substitutions (non-synonymous changes with SNP frequency >50%), respectively, between the first and last sample in the capsid (VP1) sequences compared to 1 and 6 in patients not treated with ribavirin (Patient C and Patient D).

All of the amino acid substitutions in the capsid protruding domain of Patients A and B correspond to surface exposed sites of the capsid P2 subdomain (**Figure 6.9**, **Table 6.5** and **Table 6.6**).

Figure 6.9 Patient A and B protruding (P) domain amino acid changes seen 10 and 27 days, respectively, into ribavirin therapy mapped onto GII.4 protein crystal structure (2OBT [2]); **a)** right to left rotation; **b)** top to bottom rotation. Dark Brown (top), P2 subdomain; Beige (bottom), P1 subdomain; Yellow, Dimer interface; Dark pink, HBGA binding site; pale pink, Indirect role in HBGA binding; Violet, Binding pocket predicted by Tan *et al.* 2003; Teal, Immunogenic epitopes; Green, Residue changes associated with ribavirin (dark green, in epitope; light green, not in epitope); Extra molecule, HBGA B trisaccharide. Protruding domain sequences derived from full genome sequences

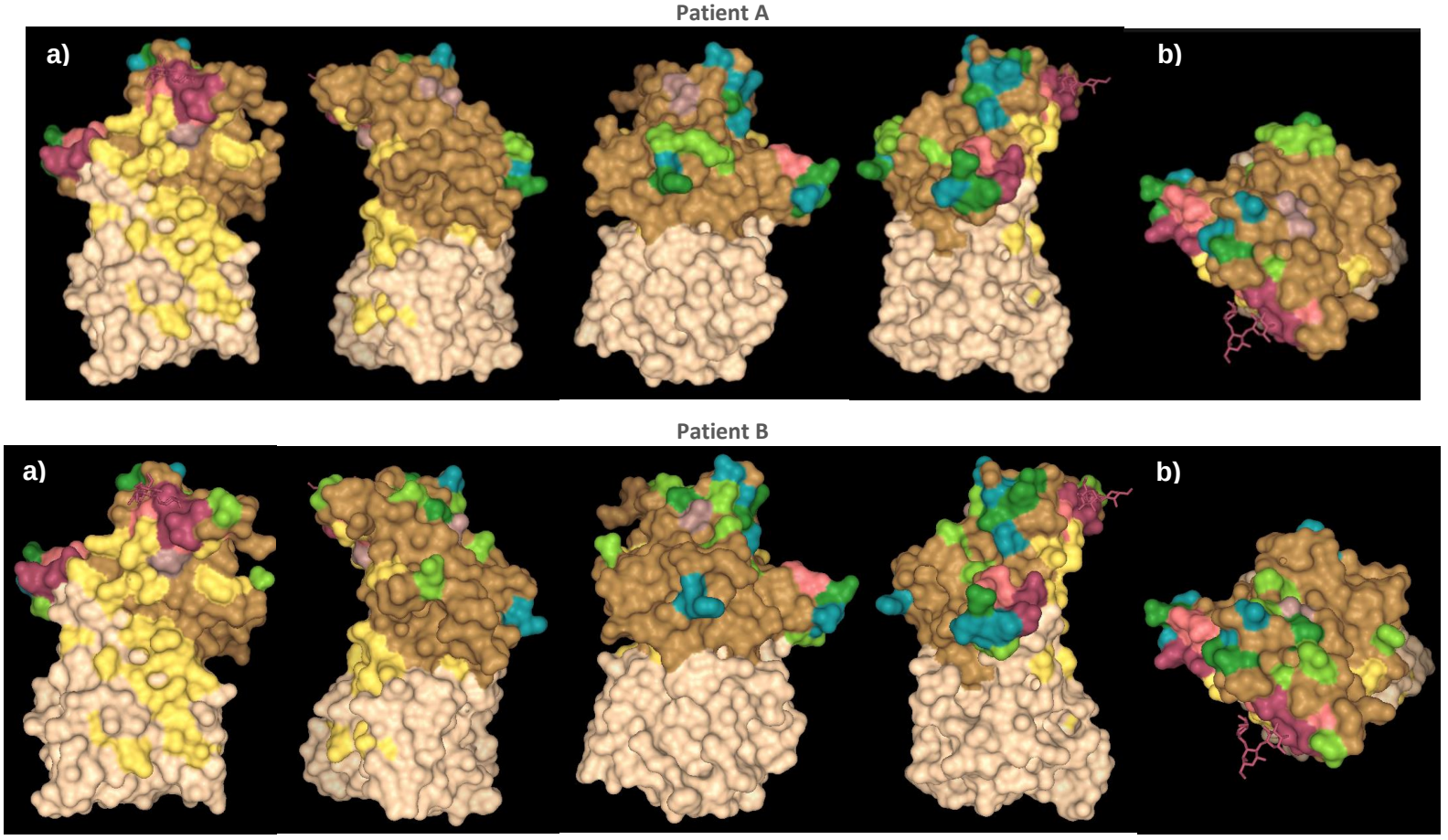


Table 6.5 Patient A capsid protruding (P) domain amino acid changes seen 10 days into ribavirin therapy (protruding domain sequences derived from full genome sequences)

Amino acid change	Known epitopes	Change to amino acid properties
Val 298 Asp	Immunogenic epitope A in P2	Hydrophobic → Acidic
Asn 372 Thr		No change (neutral)
Ser 393 Gly	Immunogenic epitope D and binding pocket indirectly involved in HBGA binding in P2	Neutral → Unique
Lys 395 Thr		Basic → Neutral
Asn 413 His	Immunogenic epitope E in P2	Neutral → Basic
	Unknown epitopes	
Phe 375 Leu	Next to HBGA binding site in P2	Hydrophobic (aromatic) → Hydrophobic (aliphatic)
Ile 410 Arg	Next to immunogenic epitope E in P2	Hydrophobic → Basic
Asn 411 Asp		Neutral → Acidic
Asp 354 Asn	Not known epitope in P2	Acidic → Neutral

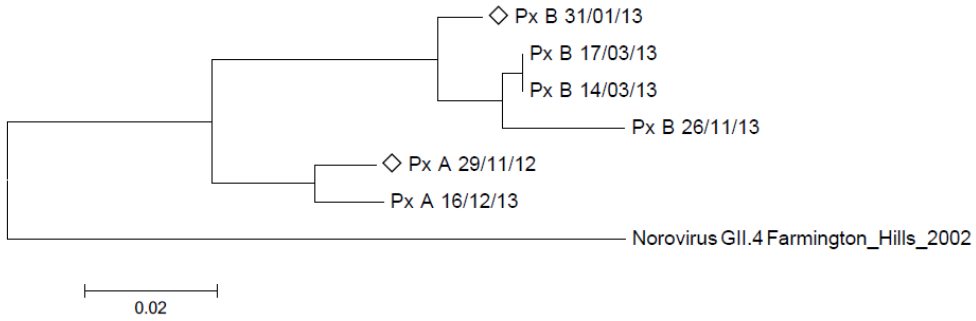
Table 6.6 Patient B capsid P domain amino acid changes seen 27 days into ribavirin therapy (protruding domain sequences derived from full genome sequences)

Amino acid change	Known epitopes	Change to amino acid properties
Thr 291 Ile	Speculative binding pocket in P2	Neutral → hydrophobic (aliphatic)
Asn 297 Arg	Immunogenic epitope A in P2	Neutral → Basic
Ser 298 Asn		No change (neutral)
Asn 372 Asp		Neutral →Acidic
Asp 393 Ser	Immunogenic epitope D and binding pocket indirectly involved in HBGA binding in P2	Acidic → Neutral
	Unknown epitopes	
Ser 310 Asn	Not known epitope in P2	No change (neutral)
Ser 352 Tyr		Serine → Hydrophobic (aromatic)
Asn 357 Asp		Neutral → acidic
Thr 293 Ile	Next to HBGA binding site in P2	Neutral → hydrophobic (aliphatic)
Asp 341 Asn		Acidic → neutral
Leu 375 Phe		Hydrophobic (aliphatic)→ hydrophobic (aromatic)
Ala 396 Pro		Hydrophobic (aliphatic) → Unique
Ala 377 Thr	Next to speculative binding pocket in P2	Hydrophobic (aliphatic) → Neutral

6.3.3.4 Phylogenetic analysis of capsid nucleotide sequences

Phylogenetic analysis of the capsid gene fragment, VP1, showed the sequence amplified from the sample from Patient B after norovirus was undetectable and diarrhoea had improved, clustered with other samples from Patient B, but in a separate cluster to samples from Patient A and a GII.4 variant Farmington Hills 2002 (**Figure 6.10**). Sequences within each cluster were 97-99.8% identical at the nucleotide level, whereas sequences between clusters, but still of the same variant type (GII.4 New Orleans_2009), had 92-93% sequence identity. GII.4 variant Farmington Hills 2002 had 85-88% sequence identity to the GII.4 New Orleans_2009 sequences.

Figure 6.10 Phylogenetic tree of Patient A and Patient B VP1 (capsid) nucleotide sequences. All sequences are genotype GII.4 variant New Orleans_2009, with the exception of the outgroup GII.4 variant Farmington Hills 2002. ◇ indicates samples collected before ribavirin therapy was started. Scale bar represents the number of substitutions per site



6.3.4 Method to determine mutation frequencies: cloning vs. deep sequencing

Each cloned amplicon generated 64–88 colony sequences for mutation frequency analysis. Deep sequencing generated a median of 835–4,763 reads per amplicon.

The mutation frequencies determined by cloning were up to ten-fold higher than by deep sequencing (**Table 6.7**), suggesting cloning considerably over-estimates the mutation frequency. Consequently deep sequencing was used to investigate the effect of ribavirin on norovirus intra-host mutation frequency.

Table 6.7 Comparison of mutation frequencies obtained by cloning and deep sequencing method

	Region	Cloning Mutation frequency per 10,000 nt	Deep Sequencing Mutation frequency per 10,000 nt
GII.4 full length clone control	NS7	15	3
	VP1	12	1
Patient A 29/11/12 (pre-ribavirin)	NS7	36	3
	VP1	35	11
Patient A 16/12/12 (during ribavirin)	NS7	32	4
	VP1	26	1

NS7, region of RNA polymerase gene; VP1, region of capsid gene

6.3.5 Mutation frequencies and ribavirin

6.3.5.1 Fold-changes in mutation frequency

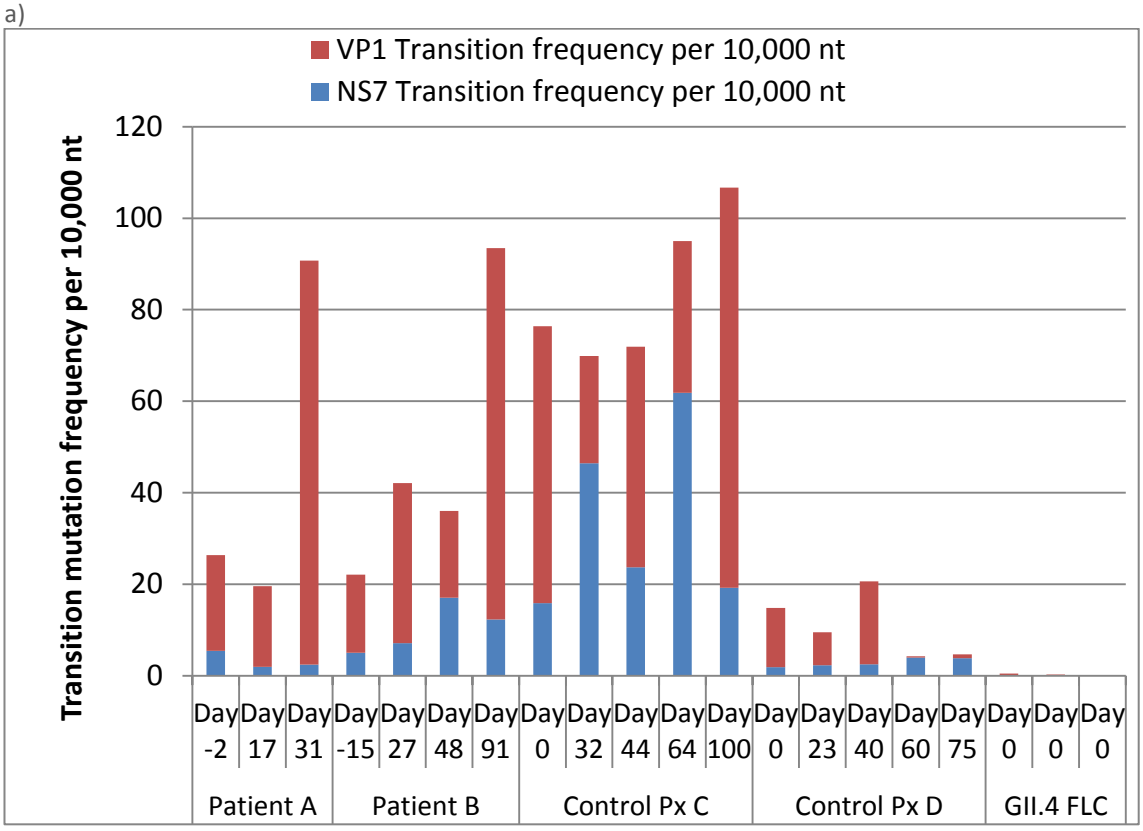
The mutation frequency of the GII.4 full length clone control was <1 mutation per 10,000 nt.

One month after start of treatment, the mutation frequency in Patient A and Patient B increased by 3.4-fold and 1.9-fold, respectively, compared to pre-ribavirin samples. During the same time period the mutation frequency in control patient C reduced by 0.9-fold and in Patient D increased 1.4-fold. Three months after the start of treatment, the mutation frequency in Patient B increased 4.2-fold compared to the pre-ribavirin sample; over the same time frame the mutation frequency in control patient C increased by 1.4-fold and in Patient D decreased 3.2-fold (**Figure 6.11**). The difference in mutation frequency between pre-and post-ribavirin samples in Patient A and Patient B was not significant (P = 1.0 and 0.5, respectively). The P values are exceptionally high due to the small sample number, in particular only one sample in the pre-ribavirin group. When the data for Patient A and Patient B is grouped together, so there are more data points in each group, the outcome is unchanged (P = 0.371).

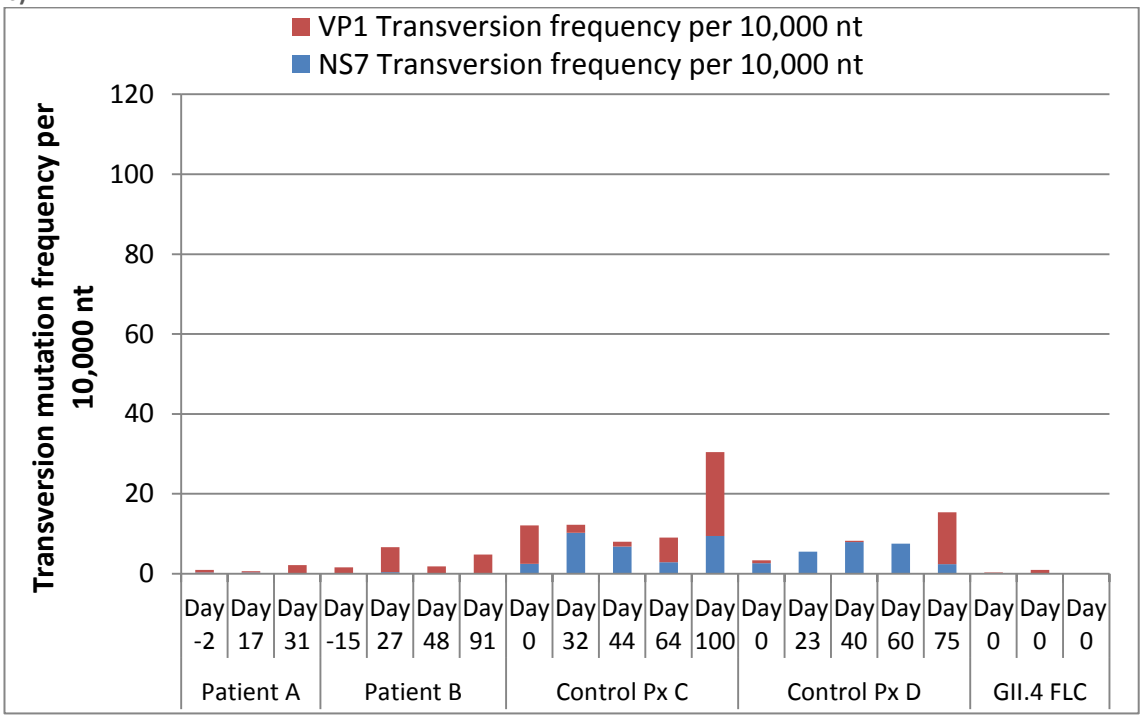
In all patients the mutation frequencies in the capsid region (median 21.0 mutations per 10,000 nt) are significantly higher than in the RdRp region (median 5.4 mutations per 10,000 nt) (P = 0.005).

Figure 6.11 Mutation frequencies for ribavirin treated Patient A and Patient B and untreated Patient C and Patient D in longitudinal samples (excluding consensus level changes). GII.4 FLC is a norovirus GII.4 (Farmington Hills) control. Day number is relative to start of ribavirin therapy. **(a)** Transitions only; **(b)** Transversions only

a)



b)



6.3.5.2 Transitions and transversions

The majority of mutations in Patients A and B were transitions, rising from 96% and 93% pre-ribavirin, respectively, to 98% and 95% during treatment. The proportion of transitions in un-treated patients C and D ranged between 78-91% and 36–81%, respectively. Therefore the proportion of transitions in Patients A and B was higher than in the un-treated patients, but this was already the case pre-ribavirin (**Figure 6.11**).

6.3.5.3 Non-synonymous and synonymous mutations (dN/dS ratios)

All patients had wavering dN/dS ratios in the capsid over time (**Table 6.8**), suggesting the selection pressure on the capsid is frequently changing.

The polymerase in Patient B and Patient C had a dN/dS ratio consistently below one (0.0–0.5). Conversely Patient D consistently had a dN/dS ratio near or above one (0.9–3.6). Patient A had a dN/dS ratio of 3.0 before the commencement of ribavirin therapy, and below one (0.5-0.7) during treatment (**Table 6.8**).

Table 6.8 Intra-host dN/dS ratios in RdRp (NS7) and capsid (VP1) in ribavirin treated Patients A and B and chronically infected un-treated Patients C and D, determined by deep sequencing of NS7 and VP1 amplicons (excluding positions with SNP frequency >50%)

		dN/dS ratio	
		NS7	**VP1**
Patient A	Day -2 (pre-ribavirin)	3.0	1.7
	Day 17 (during ribavirin)	0.7	0.3
	Day 31 (during ribavirin)	0.5	0.3
Patient B	Day -15 (pre-ribavirin)	0.0	3.2
	Day 27 (during ribavirin)	0.1	2.8
	Day 48 (during ribavirin)	0.0	2.7
	Day 91 (during ribavirin)	0.0	1.2
Patient C	Day 0	0.1	1.2
	Day 32	0.2	1.0
	Day 44	0.3	0.5
	Day 100	0.5	2.6
Patient D	Day 0	1.7	5.7
	Day 23	3.6	0.4
	Day 40	3.4	4.4
	Day 75	0.9	2.5

6.3.5.4 Frequency distribution of mutations

The distribution of low frequency mutations differs between patients. Patients A and B had non-synonymous mutations with <50% frequency at 3 and 2 positions across the NS7 amplicon sequence, whereas Patients C and D, who were not treated with ribavirin, had low frequency mutations at 15 and 12 positions (**Figure 6.12**). This difference between treated and untreated patient is not mirrored in the capsid (VP1) amplicon, for which low frequency mutations are seen at 13, 19 and 29 positions for Patients A, B and C and only 5 positions in Patient D (**Figure 6.13**).

Figure 6.12 Mutation frequency spectrum in RdRp (NS7) of longitudinal samples from ribavirin treated Patients A, B and chronically infected un-treated Patients C and D, determined by deep sequencing of NS7 amplicon. Vertical axis indicates the number of positions across the region sequenced; X axis indicates the frequency of the mutation at a given position. Day number is relative to start of ribavirin therapy

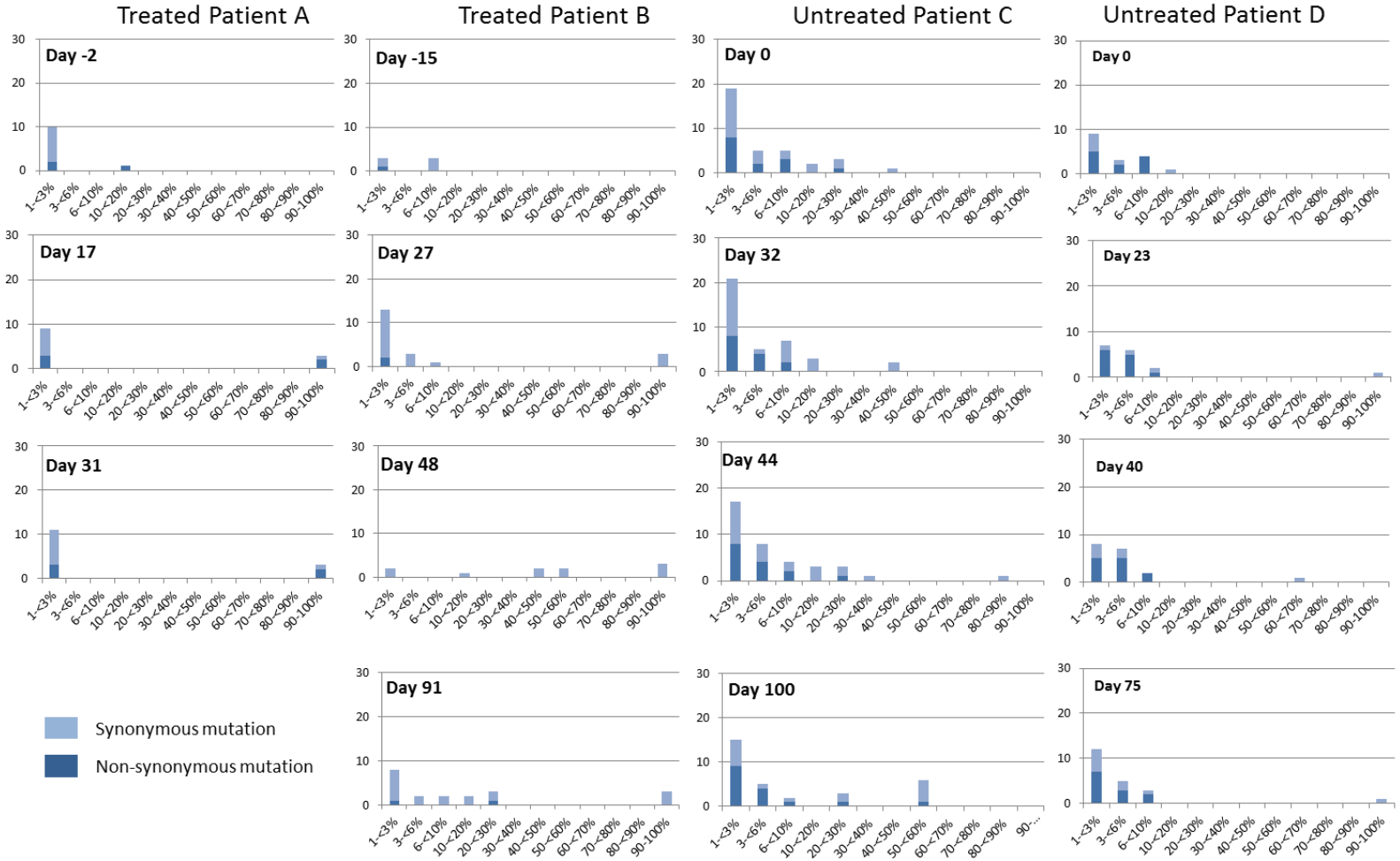
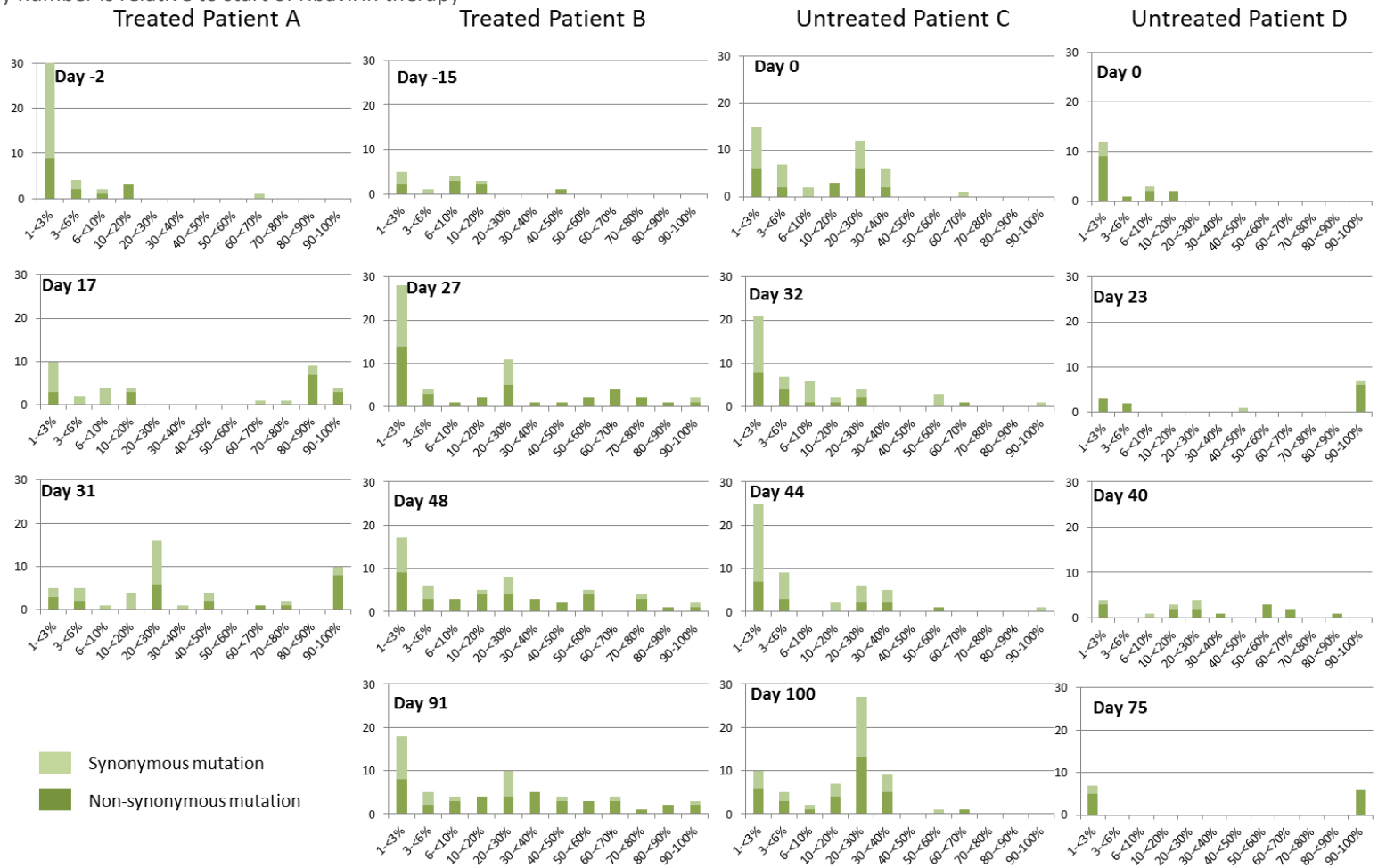


Figure 6.13 Mutation frequency spectrums in capsid (VP1) of longitudinal samples from ribavirin treated Patients A, B and chronically infected un-treated Patients C and D, determined by deep sequencing of VP1 amplicon. Vertical axis indicates the number of positions across the region sequenced; X axis indicates the frequency of the mutation at a given position. Day number is relative to start of ribavirin therapy



6.4 DISCUSSION

This chapter describes the compassionate treatment with oral ribavirin of two paediatric solid organ transplant recipients chronically infected with norovirus. Ribavirin was well tolerated by both patients, with no noted side effects or anaemia. Whilst only one of the two patients cleared norovirus during treatment both gained weight, thereby avoiding the need for parenteral nutrition. However, notwithstanding clearance of norovirus in patient A, both cases continued to have diarrhoea for several months suggesting on-going gut dysfunction following norovirus infection.

6.4.1 Amino acid substitutions in RdRp with ribavirin therapy

Patient A, who is the only patient to have cleared norovirus whilst on ribavirin therapy, is the only patient in this study to have developed amino acid substitutions in RdRp, and their appearance corresponded with a log reduction in viral load and viral clearance. The functional consequence of these mutations is unknown, however one possible mechanism is that the mutations code for a higher-fidelity RdRp which, whilst potentially minimising incorporation of ribavirin into the RNA template during replication, may also infer lower fitness therefore contributing to the elimination of norovirus in this patient. Poliovirus with a high fidelity RNA polymerase has previously been shown to have an attenuated phenotype and restricted tissue tropism [237] and high-fidelity RNA polymerase in murine norovirus, caused by a single amino acid change, has been shown to result in lower transmissibility [243]. Therefore it is plausible that reduced viral fitness in Patient A is caused by consensus-level mutations that reached fixation under selection pressure from ribavirin treatment, however functional experiments are required to assess the impact of the mutations seen in Patient A on the function of RdRp.

Whilst coding mutations were not seen in the RdRp of any other patients in this study, it has previously been described in a heart transplant recipient with chronic norovirus infection, in whom a single amino acid change was seen after 1 month of longitudinal sampling [212]. Nilsson *et al.* do not describe the position of the RdRp coding changes therefore they cannot be compared to those seen in Patient A in this study.

6.4.2 Norovirus viral loads

Whilst Patient B did appear to clear norovirus several weeks after the end of ribavirin therapy, phylogenetic analysis of the capsid confirmed that detection of the virus two months later was not a new norovirus infection, but the resurgence of the existing

infection which must have remained sequestered in the patient at an undetectable level whilst the patient was asymptomatic. Nevertheless Patient B appeared to recover gut function whilst on ribavirin therapy, which suggests attenuated virulence associated with ribavirin therapy.

As is seen in **Figure 6.6 (a)** on page 203, the reduction of viral load in Patient A is preceded by reduced immunosuppression with a lowered sirolimus dose, which could account for viral clearance due to improved immune function. However viral clearance due to reduced immunosuppression, as opposed to ribavirin, might result in a different pattern in mutation frequencies to Patient B, in whom immunosuppression (tacrolimus) was not reduced, which is not the case. Nevertheless reduced immunosuppression in Patient A may have had a synergistic effect with ribavirin to clear norovirus, which did not occur in patient B. Alternatively viral clearance in Patient A and not B could be explained by a synergistic effect between ribavirin and sirolimus, which has previously been shown to have moderate antiviral effects [225]. This is supported by the dramatic drop in viral load leading to viral clearance when the sirolimus dose was increased to 0.4 mg OD and the plasma ribavirin level reached 400 ng/ml. This was not seen in Patient B who was managed post-transplant with tacrolimus instead of sirolimus.

6.4.3 Next generation sequencing for investigation of mutation frequencies

Mutation frequencies were over-estimated by up to 10-fold when determined by cloning and capillary sequencing, compared to deep sequencing, including in the full length clone (FLC) control, which one would expect to have little or no mutations. The high error rate seen in the cloning method may be due to replication errors in *Eschericia coli* during colony growth or, most likely, the small number of sequenced clones. Although 96 colonies were capillary sequenced for each sample, only 64-88 generated high quality sequences. Even though Beaucourt *et al.* [158] suggest a minimum of 70 colonies for an accurate estimation of the mutation frequency, deep sequencing provides an average read depth of over 800 which is effectively equivalent to sequencing over 800 clones. The extra data will provide a more accurate estimate of the mutation frequency.

Deep sequencing of amplicons is also more cost effective than sequencing multiple clones. The cost of cloning, sub-culturing and sequencing 96 clones in a single direction is approximately £350 per sample, whereas the cost of paired end deep sequencing is only £57 per sample for sequencing library preparation plus £685 for a

MiSeq run, for which up to 1,700 samples can be multiplexed with 5000-fold read depth per sample.

The mutation frequency range for the GII.4 full length clone (FLC) in this study using deep sequencing is only 0–0.64 mutations per 10,000 nucleotides, suggesting that the *in vivo* mutation frequencies seen in Patients A, B, C and D are genuine, rather than PCR or sequencing artefacts.

6.4.4 Mutation frequencies are higher in the capsid

In ribavirin-treated Patients A and B the mutation frequencies in the capsid region (VP1) were significantly higher than in RdRp (NS7) and several capsid amino acid substitutions occurred at the consensus level. By contrast, no or few amino acid substitutions were seen in untreated patients sampled over a similar timeframe. All the substitutions localised to the surface exposed-sites of the capsid P2 subdomain, and may therefore represent escape mutations in response to host immune pressure. Since Patients C and D were bone marrow, rather than solid organ, transplant recipients, the immune pressures acting on the norovirus capsid may be different to Patients A and B.

The mutation frequency in RdRp was significantly lower than in the capsid for all patients. This likely reflects functional constraints on the polymerase protein; viruses with mutated polymerase genes may not be able to replicate. This is reflected in the lack of consensus level changes in the RdRp of Patients B, C and D.

6.4.5 dN/dS ratios

dN/dS ratios, obtained from deep sequencing mutation frequency data, suggest that the norovirus capsid in chronically infected patients is under frequently changing selection pressure. Conversely, distinct patterns are seen in the dN/dS ratios of the polymerase (RdRp, NS7).

In Patients B and C RdRp is under purifying selection (<1) at all times, suggesting all mutations are disadvantageous. Conversely in Patient A, who is the only patient with polymerase amino acid changes, RdRp was under positive selection (>1) pre-ribavirin but under purifying selection (<1) once on treatment. Patient D, who wasn't treated with ribavirin, had a polymerase evolving under positive selection (>1) throughout sampling. Given than Patient A's RdRp was under positive selection prior to commencement of ribavirin therapy and similar dN/dS ratios are seen in Patient D, the positive selection pressure on RdRp seen in Patient A is likely not due to ribavirin.

However a naturally occurring positive selection pressure on Patient A's RdRp may have allowed selection of a ribavirin-associated mutation. Although Patient D's RdRp appears to be under positive selection, with several low frequency coding mutations, none of these reached fixation.

6.4.6 Changes in mutation frequency with ribavirin therapy

An increase in mutation frequency is one of several mechanisms of action described for ribavirin [239] (**Table 6.2** on page 195). Whilst high mutation frequencies in RNA viruses, which lack proof-reading enzymes, provide an adaptive advantage, it has been shown in polio virus that a ribavirin-induced increase in mutation frequency beyond the error threshold is deleterious [237].

A three-fold increase in mutation frequency has been described in ribavirin treated murine norovirus *in vitro* [244], with an associated, albeit limited, reduction in infectivity; however the same effect has not been reported *in vivo* in mice infected with murine norovirus, in which the observed mutation frequencies were comparable to placebo-treated mice [244]. This is the first investigation of *in vivo* mutation frequency analysis in human norovirus infection, showing a 1.9–4.2-fold increase in mutation frequency in ribavirin-treated patients and ranging from a 3.2-fold decrease to a 1.4-fold increase in un-treated patients. This study shows an increase in mutation frequency in ribavirin treated patients beyond that seen in un-treated patients; however the baseline mutation frequencies of one un-treated patient (Patient C) were comparable to that seen in ribavirin-treated patients. Due to the small sample size the increases in mutation frequency in ribavirin-treated patients did not reach significance.

6.4.7 Suitability of untreated patients as controls

Patients C and D were chosen as controls in this study because they were immunocompromised patients chronically infected with GII.4 norovirus during the study period. However the underlying causes of immunosuppression differ; Patients A and B are solid organ transplant recipients, whereas Patients C and D are bone marrow transplant recipients therefore are likely to be more profoundly immunosuppressed.

The observed mutation frequencies for Patient C were higher than in all other patients across the majority of time points, whereas the mutation frequencies in Patient D were considerably lower than the mutation frequencies seen in all other patients. The fluctuating, and in the case of Patient C high baseline, mutation frequencies suggest that the mutation frequency in chronically infected immunocompromised patients is not constant. The rate of mutation frequencies may be a function of underlying

immunosuppression or duration of infection, although it is not known how long Patients C and D had been infected at the time of sampling.

Patient C, who had the highest baseline mutation frequency, was infected with a different ORF1 genotype compared to the other patients; Patient C's virus had a GII.Pe ORF1 whereas Patients A, B and D's viruses had a GII.P4 ORF1. It is possible that these different polymerases have different replication properties, with GII.P4 being higher fidelity and thus resulting in a lower mutation frequency compared to GII.Pe. If this were to be the case, it lends greater significance to the rise in mutation frequency seen in Patients A and B who were treated with ribavirin, as this is not seen in Patient D who is infected with the same ORF1 genotype but was not treated.

In vitro experiments, in which ribavirin has been shown to cause an increase in mutation frequency in murine norovirus [244], are controlled situations with a standardised inoculum, duration of infection, sampling time and genotype. In an observational study such as this one, on the other hand, there is variability in inoculum size, duration of infection, immune status, and even genotype, all of which may cause changes in mutation frequencies. The difference between patients makes it very difficult to draw conclusions; whilst changes in mutation frequency are seen in Patients A and B, in the context of naturally fluctuating mutation frequencies one cannot definitively say that these are caused by ribavirin therapy. For definitive conclusions, a placebo-controlled trial with controls matched for immunodeficiency cause and status, time post-infection and infecting genotype must be performed.

6.4.8 Ribavirin levels in the blood

We took a cautious approach to dosing due to concerns of toxicity in our paediatric patients; consequently neither patient sustained plasma ribavirin levels above 1000 ng/ml, which is the recommended therapeutic dose for HCV and has been associated with norovirus clearance in adult CVID patients [215]. These CVID patients remained on treatment for 6 and 14 months to achieve clearance of norovirus, which calls into question the potency of ribavirin against norovirus. However the authors did observe an inverse relationship between ribavirin plasma levels and viral load excretion in stool, suggesting ribavirin therapy was indeed responsible for viral clearance.

Since no side effects were seen in either of our patients it is likely that higher dosing would be achievable and that dosing in this study was possibly sub-optimal.

6.4.9 Other treatment options for norovirus

This is not the first attempt to treat chronic norovirus infections, although unequivocal and reproducible success is yet to be reported with any agent. Treatment with breast milk, which one would expect to contain anti-norovirus IgA, had no impact on viral load or symptoms [212, 217]. Human immunoglobulin has shown mixed results; cases treated with intravenous immunoglobulin (IVIG) have been unsuccessful [212, 214, 223, 227, 228] as have cases treated with oral immunoglobulin [212, 222, 227, 245]. The exception to treatment with oral immunoglobulin is reported by Gairard-Dory *et al.* in a case series of 12 lung transplant recipients with norovirus infection, in whom 11/12 had full symptom resolution and 1/12 mild symptom improvement when treated for 2 days with oral immunoglobulin [224]. However the norovirus status of these patients was not verified after treatment, therefore it is not known whether virus clearance accompanied clinical recovery.

The limited success of immunoglobulin treatment may be due to inconsistent levels of norovirus-specific antibodies in immunoglobulin preparations. However it could also be due to the mode of delivery, since two cases of severe chronic norovirus infection were successfully treated when immunoglobulin was delivered via the enteral route [211, 223].

Nitazoxanide, an anti-protozoal agent with additional activity against anaerobic bacteria and viruses, resolved clinical symptoms in an immunosuppressed hematopoietic stem cell transplant (HSCT) recipient with a ten day history of diarrhoea [246] and is the only agent to have been the subject of a randomized double-blind placebo controlled trial (RCT). The RCT saw a significant reduction in the duration of illness in immunocompetent patients from 2.5 days in the placebo arm to 1.5 days in the treatment arm [247]. This was a small trial, with 6–7 patients in each arm and only included immunocompetent patients with acute infections; the utility of nitazoxanide in chronic infection in immunocompromised patients is yet to be determined. In a case series of 17 norovirus infected patients in a cancer centre treated with nitazoxanide, the mean duration of diarrhoea was reduced to 7 days compared to 12 days in untreated patients, however this was not statistically significant [206]. An additional case report in a single immunocompromised patient showed that treatment with nitazoxanide for 7 days was not effective [228].

Favipiravir is a new broad range nucleoside analogue with some homology to ribavirin which is currently in Phase 3 clinical trials for the treatment of influenza in adults. As it is not yet a licenced drug it has not been used against norovirus infections in patients

but so far shows promise against murine norovirus *in vitro* and *in vivo* in mice. Compared to ribavirin, favipiravir showed a greater fold-change in mutation frequency (6-fold compared to 3-fold) in cell culture [244] and all mice infected with murine norovirus (5/5) had reduced viral titres in stool when treated with favipiravir, compared to only 2/5 mice with ribavirin. The increased mutagenic activity of favipiravir may have a greater effect on norovirus infectivity and thus may be a superior agent to ribavirin.

6.4.10 Conclusions

There is likely a complex combination of mechanisms of action involved in the *in vivo* treatment of chronic norovirus infection with ribavirin. The evidence presented in this chapter suggests it is possible that lethal mutagenesis is playing a role, however in light of differing immune pressures, durations of infection at time of treatment and fluctuating mutation frequencies also seen in untreated patients, it is difficult to make firm conclusions. There are three scenarios that may be occurring with regards to mutation frequency. First, ribavirin did not have activity against norovirus in the two patients that were treated in this study. Second, ribavirin did act directly against norovirus, but not via lethal mutagenesis. Third, ribavirin acted directly against norovirus via lethal mutagenesis, but the effect is masked by the naturally occurring changes in norovirus mutation frequency in chronic infections.

More data is required on the intra-host mutation rate in chronically infected patients and how this is affected by immune status, duration of infection and genotype. There is also a need for bigger sample sizes of ribavirin-treated patients and matched controls; ideally in the context of a randomised controlled trial or, at the very least, greater numbers of patients with comparable immune statuses.

This pilot data shows a potential role for ribavirin in treatment of chronic norovirus infections but further studies are required to increase evidence of efficacy, optimise treatment regimens and compare new alternative agents such as favipiravir.

Deep sequencing is a useful tool for analysis of mutation frequencies, with improved accuracy and cost effectiveness over conventional cloning and Sanger sequencing. At the time of specimen processing for this study norovirus full genome sequencing was not yet available; therefore this study used partial genome sequences amplified by PCR followed by deep sequencing. Future studies should assess the utility of full genome sequencing, such as described in Chapter 4, for the analysis of mutation frequencies across the norovirus genome.

6.5 ACKNOWLEDGEMENTS

Figure 6.7 was prepared by Lucy Thorne in Professor Ian Goodfellow's lab at University of Cambridge. All other preparation of figures and analysis was done by me.

CHAPTER 7

CONCLUSIONS AND DIRECTION OF FUTURE WORK

7.1 NOROVIRUS IN IMMUNOCOMPROMISED CHILDREN

The data presented in this thesis shows norovirus to be a leading cause of gastroenteritis in children, with the greatest burden falling on immunocompromised patients. Immunocompromised patients have the highest prevalence of infection compared to general medical and surgical patients, transmission clusters are dominated by immunocompromised patients which means they are involved in patient-patient transmission more frequently than medical and surgical patients (Chapter 5) and they are the most likely to develop chronic infections (Chapter 3).

Norovirus infections are associated with the highest viral titres and the highest rate of chronic infections compared to other gastrointestinal viruses (Chapter 3). It is interesting to note though that adenovirus, rotavirus and sapovirus are also associated with chronic infections, which anecdotally is seldom realised. Despite suggestions that B cells may be important for norovirus replication [24], immunocompromised children who lack B cells are equally susceptible to norovirus infection as those with B cells, albeit with a lower viral load (Chapter 3).

There is an observed difference in the prevalence of infections caused by GII.3 in children and adults (Chapter 3 and Chapter 5), however the reason for this is not known. One possibility is that the two genotypes circulate in equal numbers in children and adults in the community, but GII.4s are more transmissible and thus are seen most frequently in nosocomial outbreaks, which are dominated by adults. The data presented in this thesis does not support this hypothesis; there is no difference in the number of patients involved in transmission clusters caused by GII.3 and by GII.4 (Chapter 5). In fact, the number of patients involved in GII.3 clusters was slightly higher (38 and 33, respectively) despite observing fewer GII.3 clusters (6 clusters caused by GII.3 compared to 9 by GII.4) which suggests GII.4 is not more transmissible than GII.3.

A genuine difference in the prevalence of GII.3 in children and adults could alternatively be explained by longer lasting immunity to GII.3 or differences in the presence of GII.3 cellular receptors or co-factors between children and adults. Either possibility would result in fewer GII.3 infections in adulthood. For this to hold true one would expect to see a difference in the prevalence of GII.3 between children and adults in the community; since the distribution of norovirus genotypes in the community is not known this remains to be determined.

7.2 CHRONICALLY INFECTED PATIENTS

The majority (79%) of chronically infected patients are immunocompromised (Chapter 3). Given that some of the medical/surgical patients may also have some degree of immunosuppression, the exact proportion is probably higher.

There is no difference in the viral titre between norovirus infections in immunocompromised and immunocompetent patients, nor is there a difference in the genotypes between acute and chronic infections (Chapter 3). Both of these observations suggest the establishment of chronicity in immunocompromised patients is host, not virus, mediated.

Until now it has not been certain whether chronically infected patients are continuously infected with the same virus or whether re-infection occurs, especially in instances where there are periods of undetectable norovirus by real-time PCR. Full genome genotyping and phylogenetic analysis shows that re-infections do occur, especially following a period of undetectable real-time PCR. Nonetheless a relatively short period of undetectable norovirus (less than 2 months) can indicate persistence with the same virus (Chapter 5); in this instance the virus must persist in the stool at a very low level (below the limit of detection of PCR) or elsewhere.

Although chronically infected patients who are continuously PCR positive are most likely to remain infected with the same virus (Chapter 5) super-infections do occur, leading to a period of mixed infection that can be detected during sequence data assembly (Chapter 4). Without regular genotyping of chronically infected patients these super-infections would not be recognised. The full clinical implication of super-infection is unknown however they do not appear to increase the duration of infection (Chapter 5); this is perhaps unsurprising given that the establishment of chronicity in norovirus infections is not genotype mediated (Chapter 3). Regardless of the clinical implication to an individual patient, super-infections represent a breach of infection control.

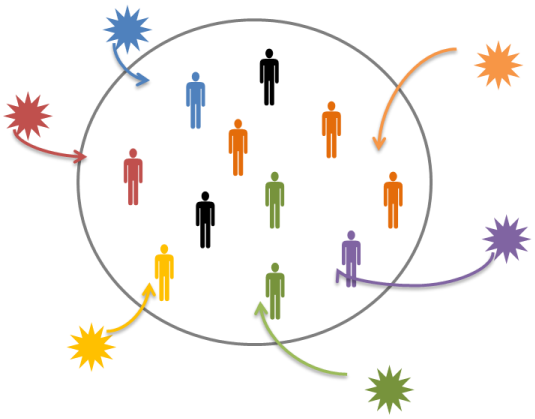
7.3 SOURCES OF INFECTION IN A PAEDIATRIC HOSPITAL

Norovirus is notorious as a cause of nosocomial outbreaks during winter, dominated by GII.4 norovirus [128]. In the community we see a breadth of genotypes causing infection [130]; although the exact distributions of genotypes in the community is unknown the diversity seen in waste water contaminated with sewage [248-250] suggests that, unlike nosocomial infections, community infections may not be dominated by GII.4.

The occurrence of norovirus infections at GOSH, a paediatric tertiary referral hospital, is not seasonal. Instead the number of new infections correlates with the number of admissions, estimated to account for 50% of the variability in the number of new infections per month (Chapter 3). Given that community norovirus outbreaks do not follow a seasonal trend [119], this begins to suggest a community source of infection.

Genotyping reveals a breadth of genotypes causing infection at GOSH, with only 39–52% of infections caused by GII.4 (Chapter 5 and Chapter 3). Furthermore full genome phylogenetic analysis suggests that 69% of new norovirus infections are not acquired from another patient (Chapter 5). This points to a model of transmission in which there are frequent introductions of multiple strains, with limited onward transmission of each (**Figure 7.1**).

Figure 7.1 Proposed transmission dynamics at paediatric tertiary referral hospital (GOSH), with multiple introductions of several different strains with limited transmission of each



Combined with the lack of seasonality, the breadth of genotypes at GOSH suggests a community source of infection. A substantial proportion of this (45%) represents patients who were norovirus positive on admission (POA) thus represent infection prior to hospital admission, most likely whilst in the community or at another healthcare facility. Of the remaining 54% which were hospital acquired infections (HAI), the distribution of genotypes more closely resembles community norovirus infections (Chapter 3, Chapter 5). Half of the HAIs (57/101) are explained by nosocomial patient-patient transmission; however the remaining HAIs, making up a quarter of all norovirus infections during the study period, have an unknown nosocomial source. This begs the question, what is the source of these infections?

One explanation is transmission from staff with norovirus infections, since their infection will have been acquired in the community. However staff are unlikely to be the

main source of infection since they are present in all healthcare institutions, yet the epidemiology seen at GOSH differs to other healthcare settings (Chapter 5).

Norovirus is an extremely diverse species, yet the reservoir in which the broad genotypes are maintained and from which new epidemic and pandemic strains arise is currently not known. Since there is no evidence of zoonotic transmission [16] an animal reservoir is deemed unlikely; current opinion is that the reservoir of norovirus diversity is within the human population [251]. The high level of intra-host norovirus diversity reported in immunocompromised patients, but not seen in immunocompetent infections, has led to suggestions that immunocompromised people may be a reservoir for norovirus diversity [52]. In particular a weak immune response, which fails to completely shut down viral replication (as in an immunocompetent person) but simultaneously provides selective pressure for escape mutants (unlike in profoundly immunocompromised people), is postulated to result in maximal intra-host virus adaptation and thus a reservoir of diversity [252]. An extension of this is the suggestion that malnourished and elderly individuals, with consequent weakened immune response, are a potential reservoir [251].

Another plausible reservoir for viral mixing is children in the community, who are estimated to experience four episodes of diarrhoeal disease per year [106] and in whom over 20% of under-five year olds experience sub-clinical norovirus infections [83]. Due to a maturing immune system, the immune response to a pathogen is weaker in a child than in an adult [253]. This comparatively weak immune response could maximise intra-host virus adaptation, creating a reservoir for norovirus infection and diversity.

GOSH is a paediatric hospital, therefore the vast majority of visitors are patients' parents, who may have other children at home or also visiting their sibling. If children in the community are a reservoir of norovirus infections and diversity they are a probable source of infection in a paediatric hospital such as GOSH, either directly or via their parents, thus accounting for the breadth of genotypes causing infection. Good hand hygiene is critical for control of norovirus [88], however adherence to hand hygiene in hospital visitors is estimated to be extremely low [254-256]. Even when alcohol hand sanitizers are placed in convenient locations in hospital lobbies, only 7% of visitors are observed using them when entering a hospital [254]. Randle *et al.* investigated hand hygiene (either with alcohol sanitizer or soap and water) on two paediatric wards; one oncology ward and one general medical/respiratory ward. Whilst hand hygiene rates were higher than was observed in hospital lobbies, they reported that only 38% of visitors adhered to hand hygiene, compared to 74–84% of healthcare professionals

[257]. This data, combined with less awareness of personal hygiene in children [258], supports parents and children as plausible vectors of infection in a paediatric hospital.

Indirect transmission of norovirus, most likely via staff, visitors or contaminated fomites, plays an important role in norovirus nosocomial transmission (Chapter 5). To definitively identify the routes of indirect transmission sampling of staff, the environment and visitors is required, although this is unlikely to be practicable on a routine basis. Instead infection control efforts should be focused on hand hygiene by staff and visitors moving between wards but also careful consideration should be given to the movement of portable objects and equipment between wards.

7.4 FULL GENOME SEQUENCING IN CLINICAL PRACTICE

The advancement of sequencing techniques, from PCR with capillary sequencing to target enrichment with deep sequencing, has facilitated the use of norovirus full genomes in clinical practice. In conjunction with growing expertise, lower costs and faster turn-around times, full genomes can be sequenced for under £100 in less than a week (Chapter 4); this makes full genome sequencing a reality not just in academic settings but for informing public health practice in real time. The utility of generating full genome sequences is demonstrated in Chapter 5, in particular to identify transmission between patients who are separated by time, are on different wards, or who have not had direct contact, such as out-patients.

The requirement for full genome sequencing is limited to immunocompromised populations, as in immunocompetent populations sequences between affected patients are conserved enough to allow phylogenetic analysis using partial sequences, namely the capsid P2 domain (Chapter 5).

Despite the potential for affordable sequencing of full genomes, low sequencing costs can only be achieved by sequencing large batches of samples (48–96 samples) at a time. This is not feasible for individual hospitals that do not see such high numbers of new infections simultaneously; instead batch sequencing is better suited to a regional service. The limitation of a regional service is that turn-around times may be extended due to specimen transport time. Even so, the short incubation time for norovirus (12-48 hours) means that sequencing, which at best takes 6 days, is unlikely to be informative in the context of an acute outbreak. Results from whole genome sequencing, or simply genotyping for that matter, can be acted upon in the context of an on-going uncontrolled outbreak to identify unrecognised routes and sources of infection. In the context of a fast-spreading outbreak the results may not be timely enough, however they will still be useful to retrospectively identify outbreaks. This retrospective

information will be useful to better focus infection control resources and practices and thus prevent future outbreaks from occurring.

7.5 TREATMENT OF CHRONIC INFECTIONS

Anti-viral treatment for norovirus has not been a priority due to the self-limiting nature of most infections, typically with limited sequelae. However immunosuppressed patients are at risk of chronic infections (Chapter 3) with significant associated morbidity (Chapter 6), therefore the development of effective antiviral treatment is important for this patient population. Reports to date of treatment with various agents have shown mixed success and often small patient numbers. Often these are case series reports, rather than randomised controlled trials (with the exception of nitazoxanide) therefore it is difficult to determine whether the treatment successes are directly due to the agent that was used. The use of ribavirin for treatment of two chronic norovirus infections presented in this thesis (Chapter 6) presents a similar picture; whilst it resulted in symptom resolution in both patients and viral clearance in one, it is difficult to ascertain whether this was directly due to the action of ribavirin. Deep sequencing data could be a useful technique to determine a direct effect due to lethal mutagenesis; however the small number of patients and fluctuating mutation frequencies in untreated patents makes the results difficult to interpret. Favipiravir appears to be a more promising candidate for the treatment of chronic norovirus infections, but the effectiveness in human infections is yet to be determined.

Vaccination of vulnerable patients could be an effective way to prevent chronic infections in immunocompromised patients. Unlike the development of a specific antiviral agent, for which there is limited economic incentive due to the relatively small number of patients affected by chronic infections, a norovirus vaccine can also be used in immunocompetent people to reduce the global burden of acute infectious gastroenteritis as has been done successfully for rotavirus [259]. However vaccine development is hampered by short-lived immunity and antigenic variability. Current vaccine candidates only contain two genotypes, GI.1 and GII.4 [101]. Published *in vitro* data suggests these vaccine candidates elicit the production of cross-reactive antibodies, however the data presented in this thesis suggests this may not be the case *in vivo* in immunocompromised hosts. Patients for whom longitudinal norovirus positive samples were sequenced showed that whilst re-infection with the same genotype was not detected during the sampling period (with the exception of one patient with over a year in between infections), re-infection and super-infection with a different genotype did occur (Chapter 5). This suggests the immune response raised by

a norovirus infection may not be cross-protective across genotypes in this patient population; given the breadth of genotypes seen in a paediatric population (Chapter 3 and Chapter 5) this is likely to be problematic for the control of norovirus in immunosuppressed patients; the extent to which this applies to immunocompetent hosts, and therefore the control of norovirus globally, is yet to be determined. Future vaccine candidates will require careful consideration of the choice and number of genotypes to be included, as well as the immune status of vaccine recipients.

7.6 DIRECTION OF FUTURE WORK

7.6.1 Genotyping of community infections

Current estimates of diversity in community acquired norovirus infections are based on patients presenting to healthcare facilities, therefore have a presentation bias. To determine the role of community infections in norovirus transmission and nosocomial epidemiology, unbiased estimates of circulating genotypes are required. This could be achieved by genotyping archived norovirus positive stool samples from the Infectious Intestinal Disease Study 2 (IID2), a prospective population-based study in which people in the community were sampled longitudinally [113], or the MAL-ED study multi-country birth cohort with two-year longitudinal sampling [175]. The only limitation of the IID2 study is that samples were collected in 2008 and 2009 therefore the circulating genotypes will likely have changed. Nevertheless the genotypes identified could be compared to retrospective data from Public Health England surveillance of nosocomial outbreaks during the same time period.

7.6.2 Model for identification of transmission

There are several factors to consider when investigating possible nosocomial transmission of norovirus; the relatedness of norovirus genome sequences from each patient (evaluated by phylogenetic analysis) and the in-patient histories of each patient (Chapter 5). The in-patient histories are important; if there is no plausible link between two patients, either in time or by location, then it may be that they are coincidently infected with the same virus rather than a transmission event having occurred. Conversely indirect transmission must be considered, such as attendance at out-patient clinics or shared clinical teams across wards. A model that incorporates all of these pieces of information could be used by infection control teams to ascertain the likelihood of transmission between two patients having occurred and so is the focus of future work.

7.6.3 Chronic Viral Gastroenteritis study

A study is currently underway at GOSH called the Chronic Viral Gastroenteritis Study, set up by myself and approved by the Brent NRES committee. The aim of the study is to assess the gut pathology in acute and chronic norovirus infections, including factors that may predispose patients to chronic infections. The study has recruited 59 patients undergoing bone marrow or solid organ transplant. Each patient has stool and urine specimens collected pre-transplant and monthly for one year following transplant. Gut inflammation, gut epithelial cell damage and stool microbial community diversity will be assessed to determine how these differ in acute and chronic norovirus infections.

References

1. Zahorsky J. Hyperemesis hiemis or the winter vomiting disease. Archives of Paediatrics **1929**; 46: 391-5.

2. Kapikian AZ, Wyatt RG, Dolin R, Thornhill TS, Kalica AR, Chanock RM. Visualization by immune electron microscopy of a 27-nm particle associated with acute infectious nonbacterial gastroenteritis. Journal of virology **1972**; 10(5): 1075-81.

3. Mayo MA. A summary of taxonomic changes recently approved by ICTV. Archives of virology **2002**; 147(8): 1655-63.

4. Green K. Caliciviridae: the noroviruses. In: Knipe DM HP, Cohen JI, Griffin DE, Lamb RA, Martin MA, Racaniello VR, Roizman B. Fields Virology. 6th ed Vol. 1. Philadelphia, PA: Lippincott Williams & Wilkins, **2013**.

5. Jiang X, Wang M, Wang K, Estes MK. Sequence and genomic organization of Norwalk virus. Virology **1993**; 195(1): 51-61.

6. Xi J, Graham D, Wang K, Estes M. Norwalk virus genome cloning and characterization. Science (New York, NY) **1990**; 250(4987): 1580-3.

7. Thorne LG, Goodfellow IG. Norovirus gene expression and replication. The Journal of general virology **2014**; 95(Pt 2): 278-91.

8. Vongpunsawad S, Venkataram Prasad BV, Estes MK. Norwalk Virus Minor Capsid Protein VP2 Associates within the VP1 Shell Domain. Journal of virology **2013**; 87(9): 4818-25.

9. Pinto P, Wang Q, Chen N, et al. Discovery and genomic characterization of noroviruses from a gastroenteritis outbreak in domestic cats in the US. PloS one **2012**; 7(2): e32739.

10. Sugieda M, Nakajima S. Viruses detected in the caecum contents of healthy pigs representing a new genetic cluster in genogroup II of the genus "Norwalk-like viruses". Virus research **2002**; 87(2): 165-72.

11. Wolf S, Williamson W, Hewitt J, et al. Molecular detection of norovirus in sheep and pigs in New Zealand farms. Veterinary microbiology **2009**; 133(1-2): 184-9.

12. Karst SM, Wobus CE, Lay M, Davidson J, Virgin HWt. STAT1-dependent innate immunity to a Norwalk-like virus. Science (New York, NY) **2003**; 299(5612): 1575-8.

13. Oliver SL, Dastjerdi AM, Wong S, et al. Molecular characterization of bovine enteric caliciviruses: a distinct third genogroup of noroviruses (Norwalk-like viruses) unlikely to be of risk to humans. Journal of virology **2003**; 77(4): 2789-98.

14. Tse H, Chan WM, Lam CS, Lau SK, Woo PC, Yuen KY. Complete genome sequences of novel rat noroviruses in Hong Kong. Journal of virology **2012**; 86(22): 12435-6.

15. Martella V, Lorusso E, Decaro N, et al. Detection and molecular characterization of a canine norovirus. Emerging infectious diseases **2008**; 14(8): 1306-8.

16. Wilhelm B, Waddell L, Greig J, Rajic A, Houde A, McEwen SA. A scoping review of the evidence for public health risks of three emerging potentially zoonotic viruses: hepatitis E virus, norovirus, and rotavirus. Preventive veterinary medicine **2015**; 119(1-2): 61-79.

17. Bank-Wolf BR, Konig M, Thiel HJ. Zoonotic aspects of infections with noroviruses and sapoviruses. Veterinary microbiology **2010**; 140(3-4): 204-12.

18. Kroneman A, Vega E, Vennema H, et al. Proposal for a unified norovirus nomenclature and genotyping. Archives of virology **2013**; 158(10): 2059-68.

19. Bull RA, Hansman GS, Clancy LE, Tanaka MM, Rawlinson WD, White PA. Norovirus recombination in ORF1/ORF2 overlap. Emerging infectious diseases **2005**; 11(7): 1079-85.

20. Bull RA, Tanaka MM, White PA. Norovirus recombination. The Journal of general virology **2007**; 88(Pt 12): 3347-59.

21. Steinhauer DA, Domingo E, Holland JJ. Lack of evidence for proofreading mechanisms associated with an RNA virus polymerase. Gene **1992**; 122(2): 281-8.

22. Smith DB, McAllister J, Casino C, Simmonds P. Virus 'quasispecies': making a mountain out of a molehill? The Journal of general virology **1997**; 78 (Pt 7): 1511-9.

23. Duizer E, Schwab KJ, Neill FH, Atmar RL, Koopmans MP, Estes MK. Laboratory efforts to cultivate noroviruses. The Journal of general virology **2004**; 85(Pt 1): 79-87.

24. Jones MK, Watanabe M, Zhu S, et al. Enteric bacteria promote human and mouse norovirus infection of B cells. Science (New York, NY) **2014**; 346(6210): 755-9.

25. Jones MK, Grau KR, Costantini V, et al. Human norovirus culture in B cells. Nature protocols **2015**; 10(12): 1939-47.

26. Wobus CE, Karst SM, Thackray LB, et al. Replication of Norovirus in cell culture reveals a tropism for dendritic cells and macrophages. PLoS biology **2004**; 2(12): e432.

27. Herbst-Kralovetz MM, Radtke AL, Lay MK, et al. Lack of norovirus replication and histo-blood group antigen expression in 3-dimensional intestinal epithelial cells. Emerging infectious diseases **2013**; 19(3): 431-8.

28. Papafragkou E, Hewitt J, Park GW, Greening G, Vinje J. Challenges of culturing human norovirus in three-dimensional organoid intestinal cell culture models. PloS one **2014**; 8(6): e63485.

29. Takanashi S, Saif LJ, Hughes JH, et al. Failure of propagation of human norovirus in intestinal epithelial cells with microvilli grown in three-dimensional cultures. Archives of virology **2014**; 159(2): 257-66.

30. Straub TM, Honer zu Bentrup K, Orosz-Coghlan P, et al. In vitro cell culture infectivity assay for human noroviruses. Emerging infectious diseases **2007**; 13(3): 396-403.

31. Dolin R, Levy AG, Wyatt RG, Thornhill TS, Gardner JD. Viral gastroenteritis induced by the Hawaii agent. Jejunal histopathology and serologic response. The American journal of medicine **1975**; 59(6): 761-8.

32. Agus SG, Dolin R, Wyatt RG, Tousimis AJ, Northrup RS. Acute infectious nonbacterial gastroenteritis: intestinal histopathology. Histologic and enzymatic alterations during illness produced by the Norwalk agent in man. Annals of internal medicine **1973**; 79(1): 18-25.

33. Karandikar UC, Crawford SE, Ajami NJ, et al. Detection of human norovirus in intestinal biopsies from immunocompromised transplant patients. The Journal of general virology **2016**.

34. Ettayebi K, Crawford SE, Murakami K, et al. Replication of human noroviruses in stem cell-derived human enteroids. Science (New York, NY) **2016**; 353(6306): 1387-93.

35. Cao S, Lou Z, Tan M, et al. Structural basis for the recognition of blood group trisaccharides by norovirus. Journal of virology **2007**; 81(11): 5949-57.

36. Marionneau S, Cailleau-Thomas A, Rocher J, et al. ABH and Lewis histo-blood group antigens, a model for the meaning of oligosaccharide diversity in the face of a changing world. Biochimie **2001**; 83(7): 565-73.

37. Ferrer-Admetlla A, Sikora M, Laayouni H, et al. A natural history of FUT2 polymorphism in humans. Molecular biology and evolution **2009**; 26(9): 1993-2003.

38. Lindesmith L, Moe C, Marionneau S, et al. Human susceptibility and resistance to Norwalk virus infection. Nature medicine **2003**; 9(5): 548-53.

39. Frenck R, Bernstein DI, Xia M, et al. Predicting susceptibility to norovirus GII.4 by use of a challenge model involving humans. The Journal of infectious diseases **2012**; 206(9): 1386-93.

40. Parrino TA, Schreiber DS, Trier JS, Kapikian AZ, Blacklow NR. Clinical immunity in acute gastroenteritis caused by Norwalk agent. The New England journal of medicine **1977**; 297(2): 86-9.

41. Currier RL, Payne DC, Staat MA, et al. Innate Susceptibility to Norovirus Infections Influenced by FUT2 Genotype in a United States Pediatric Population. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2015**; 60(11): 1631-8.

42. Nordgren J, Kindberg E, Lindgren PE, Matussek A, Svensson L. Norovirus gastroenteritis outbreak with a secretor-independent susceptibility pattern, Sweden. Emerging infectious diseases **2010**; 16(1): 81-7.

43. Van Trang N, Vu HT, Le NT, Huang P, Jiang X, Anh DD. Association between norovirus and rotavirus infection and histo-blood group antigen types in Vietnamese children. Journal of clinical microbiology **2014**; 52(5): 1366-74.

44. Harrington PR, Vinje J, Moe CL, Baric RS. Norovirus capture with histo-blood group antigens reveals novel virus-ligand interactions. Journal of virology **2004**; 78(6): 3035-45.

45. Springer GF, Williamson P, Brandes WC. BLOOD GROUP ACTIVITY OF GRAM-NEGATIVE BACTERIA. The Journal of experimental medicine **1961**; 113(6): 1077-93.

46. Rasko DA, Wang G, Monteiro MA, Palcic MM, Taylor DE. Synthesis of mono- and di-fucosylated type I Lewis blood group antigens by Helicobacter pylori. European journal of biochemistry **2000**; 267(19): 6059-66.

47. Yi W, Shao J, Zhu L, et al. Escherichia coli O86 O-antigen biosynthetic gene cluster and stepwise enzymatic synthesis of human blood group B antigen tetrasaccharide. Journal of the American Chemical Society **2005**; 127(7): 2040-1.

48. Baldridge MT, Nice TJ, McCune BT, et al. Commensal microbes and interferon-lambda determine persistence of enteric murine norovirus infection. Science (New York, NY) **2014**.

49. Karst SM, Zhu S, Goodfellow IG. The molecular pathology of noroviruses. The Journal of pathology **2015**; 235(2): 206-16.

50. Chang KO, Sosnovtsev SV, Belliot G, King AD, Green KY. Stable expression of a Norwalk virus RNA replicon in a human hepatoma cell line. Virology **2006**; 353(2): 463-73.

51. Domingo E, Escarmis C, Sevilla N, et al. Basic concepts in RNA virus evolution. FASEB journal : official publication of the Federation of American Societies for Experimental Biology **1996**; 10(8): 859-64.

52. Bull RA, Eden JS, Luciani F, McElroy K, Rawlinson WD, White PA. Contribution of intra- and interhost dynamics to norovirus evolution. Journal of virology **2012**; 86(6): 3219-29.

53. Sukhrie FH, Siebenga JJ, Beersma MF, Koopmans M. Chronic shedders as reservoir for nosocomial transmission of norovirus. Journal of clinical microbiology **2010**; 48(11): 4303-5.

54. Saif MA, Bonney DK, Bigger B, et al. Chronic norovirus infection in pediatric hematopoietic stem cell transplant recipients: a cause of prolonged intestinal failure requiring intensive nutritional support. Pediatric transplantation **2011**; 15(5): 505-9.

55. Chachu KA, LoBue AD, Strong DW, Baric RS, Virgin HW. Immune mechanisms responsible for vaccination against and clearance of mucosal and lymphatic norovirus infection. PLoS pathogens **2008**; 4(12): e1000236.

56. Tacket CO, Sztein MB, Losonsky GA, Wasserman SS, Estes MK. Humoral, mucosal, and cellular immune responses to oral Norwalk virus-like particles in volunteers. Clinical immunology (Orlando, Fla) **2003**; 108(3): 241-7.

57. Chachu KA, Strong DW, LoBue AD, Wobus CE, Baric RS, Virgin HWt. Antibody is critical for the clearance of murine norovirus infection. Journal of virology **2008**; 82(13): 6610-7.

58. Karst SM. Pathogenesis of noroviruses, emerging RNA viruses. Viruses **2010**; 2(3): 748-81.

59. Atmar RL, Opekun AR, Gilger MA, et al. Norwalk virus shedding after experimental human infection. Emerging infectious diseases **2008**; 14(10): 1553-7.

60. Johnson PC, Mathewson JJ, DuPont HL, Greenberg HB. Multiple-challenge study of host susceptibility to Norwalk gastroenteritis in US adults. The Journal of infectious diseases **1990**; 161(1): 18-21.

61. Wyatt RG, Dolin R, Blacklow NR, et al. Comparison of three agents of acute infectious nonbacterial gastroenteritis by cross-challenge in volunteers. The Journal of infectious diseases **1974**; 129(6): 709-14.

62. Simmons K, Gambhir M, Leon J, Lopman B. Duration of immunity to norovirus gastroenteritis. Emerging infectious diseases **2013**; 19(8): 1260-7.

63. Kroneman A, Verhoef L, Harris J, et al. Analysis of integrated virological and epidemiological reports of norovirus outbreaks collected within the Foodborne Viruses in Europe network from 1 July 2001 to 30 June 2006. Journal of clinical microbiology **2008**; 46(9): 2959-65.

64. Kirby AE, Streby A, Moe CL. Vomiting as a Symptom and Transmission Risk in Norovirus Illness: Evidence from Human Challenge Studies. PloS one **2016**; 11(4): e0143759.

65. Teunis PF, Moe CL, Liu P, et al. Norwalk virus: how infectious is it? Journal of medical virology **2008**; 80(8): 1468-76.

66. Milbrath MO, Spicknall IH, Zelner JL, Moe CL, Eisenberg JN. Heterogeneity in norovirus shedding duration affects community risk. Epidemiology and infection **2013**; 141(8): 1572-84.

67. Sukhrie FH, Teunis P, Vennema H, et al. Nosocomial transmission of norovirus is mainly caused by symptomatic cases. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2012**; 54(7): 931-7.

68. Chen CC, Huang JL, Chang CJ, Kong MS. Fecal calprotectin as a correlative marker in clinical severity of infectious diarrhea and usefulness in evaluating bacterial or viral pathogens in children. Journal of pediatric gastroenterology and nutrition **2012**; 55(5): 541-7.

69. Marshall JK, Thabane M, Borgaonkar MR, James C. Postinfectious irritable bowel syndrome after a food-borne outbreak of acute gastroenteritis attributed to a viral pathogen. Clinical gastroenterology and hepatology : the official clinical practice journal of the American Gastroenterological Association **2007**; 5(4): 457-60.

70. Khan RR, Lawson AD, Minnich LL, et al. Gastrointestinal norovirus infection associated with exacerbation of inflammatory bowel disease. Journal of pediatric gastroenterology and nutrition **2009**; 48(3): 328-33.

71. Turcios-Ruiz RM, Axelrod P, St John K, et al. Outbreak of necrotizing enterocolitis caused by norovirus in a neonatal intensive care unit. The Journal of pediatrics **2008**; 153(3): 339-44.

72. Stuart RL, Tan K, Mahar JE, et al. An outbreak of necrotizing enterocolitis associated with norovirus genotype GII.3. The Pediatric infectious disease journal **2010**; 29(7): 644-7.

73. Pelizzo G, Nakib G, Goruppi I, et al. Isolated colon ischemia with norovirus infection in preterm babies: a case series. Journal of medical case reports **2013**; 7: 108.

74. Kawano G, Oshige K, Syutou S, et al. Benign infantile convulsions associated with mild gastroenteritis: a retrospective study of 39 cases including virological tests and efficacy of anticonvulsants. Brain & development **2007**; 29(10): 617-22.

75. Chen SY, Tsai CN, Lai MW, et al. Norovirus infection as a cause of diarrhea-associated benign infantile seizures. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2009**; 48(7): 849-55.

76. Sanchez-Fauquier A, Gonzalez-Galan V, Arroyo S, Roda D, Pons M, Garcia JJ. Norovirus-associated encephalitis in a previously healthy 2-year-old girl. The Pediatric infectious disease journal **2015**; 34(2): 222-3.

77. Ito S, Takeshita S, Nezu A, et al. Norovirus-associated encephalopathy. The Pediatric infectious disease journal **2006**; 25(7): 651-2.

78. Green KY. Norovirus infection in immunocompromised hosts. Clin Microbiol Infect **2014**; 20(8): 717-23.

79. Bok K, Green KY. Norovirus gastroenteritis in immunocompromised patients. The New England journal of medicine **2012**; 367(22): 2126-32.

80. Schwartz S, Vergoulidou M, Schreier E, et al. Norovirus gastroenteritis causes severe and lethal complications after chemotherapy and hematopoietic stem cell transplantation. Blood **2011**; 117(22): 5850-6.

81. Roos-Weil D, Ambert-Balay K, Lanternier F, et al. Impact of norovirus/sapovirus-related diarrhea in renal transplant recipients hospitalized for diarrhea. Transplantation **2011**; 92(1): 61-9.

82. Ludwig A, Adams O, Laws HJ, Schroten H, Tenenbaum T. Quantitative detection of norovirus excretion in pediatric patients with cancer and prolonged gastroenteritis and shedding of norovirus. Journal of medical virology **2008**; 80(8): 1461-7.

83. Phillips G, Tam CC, Rodrigues LC, Lopman B. Prevalence and characteristics of asymptomatic norovirus infection in the community in England. Epidemiology and infection **2010**; 138(10): 1454-8.

84. Frange P, Touzot F, Debre M, et al. Prevalence and clinical impact of norovirus fecal shedding in children with inherited immune deficiencies. The Journal of infectious diseases **2012**; 206(8): 1269-74.

85. Teunis PF, Sukhrie FH, Vennema H, Bogerman J, Beersma MF, Koopmans MP. Shedding of norovirus in symptomatic and asymptomatic infections. Epidemiology and infection **2015**; 143(8): 1710-7.

86. Ambert-Balay K, Pothier P. Evaluation of 4 immunochromatographic tests for rapid detection of norovirus in faecal samples. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology **2013**; 56(3): 194-8.

87. Costantini V, Grenz L, Fritzinger A, et al. Diagnostic accuracy and analytical sensitivity of IDEIA Norovirus assay for routine screening of human norovirus. Journal of clinical microbiology **2010**; 48(8): 2770-8.

88. NorovirusWorkingParty. Guidelines for the management of norovirus outbreaks in acute and community health and social care settings. Available at: https://www.gov.uk/government/publications/norovirus-managing-outbreaks-in-acute-and-community-health-and-social-care-settings.

89. Park GW, Barclay L, Macinga D, Charbonneau D, Pettigrew CA, Vinje J. Comparative efficacy of seven hand sanitizers against murine norovirus, feline calicivirus, and GII.4 norovirus. Journal of food protection **2010**; 73(12): 2232-8.

90. Macinga DR, Sattar SA, Jaykus LA, Arbogast JW. Improved inactivation of nonenveloped enteric viruses and their surrogates by a novel alcohol-based hand sanitizer. Applied and environmental microbiology **2008**; 74(16): 5047-52.

91. Bolton SL, Kotwal G, Harrison MA, Law SE, Harrison JA, Cannon JL. Sanitizer efficacy against murine norovirus, a surrogate for human norovirus, on stainless steel surfaces when using three application methods. Applied and environmental microbiology **2013**; 79(4): 1368-77.

92. Sickbert-Bennett EE, Weber DJ, Gergen-Teague MF, Sobsey MD, Samsa GP, Rutala WA. Comparative efficacy of hand hygiene agents in the reduction of bacteria and viruses. American journal of infection control **2005**; 33(2): 67-77.

93. Seitz SR, Leon JS, Schwab KJ, et al. Norovirus infectivity in humans and persistence in water. Applied and environmental microbiology **2011**; 77(19): 6884-8.

94. Barclay L, Park GW, Vega E, et al. Infection control for norovirus. Clin Microbiol Infect **2014**; 20(8): 731-40.

95. Morter S, Bennet G, Fish J, et al. Norovirus in the hospital setting: virus introduction and spread within the hospital environment. The Journal of hospital infection **2011**; 77(2): 106-12.

96. Cummins M, Ready D. Role of the Hospital Environment in Norovirus Containment. The Journal of infectious diseases **2016**; 213 Suppl 1: S12-4.

97. Vivancos R, Sundkvist T, Barker D, Burton J, Nair P. Effect of exclusion policy on the control of outbreaks of suspected viral gastroenteritis: Analysis of outbreak investigations in care homes. American journal of infection control **2010**; 38(2): 139-43.

98. PHE. Suspected and laboratory-confirmed reported norovirus outbreaks in hospitals. Available at: https://www.gov.uk/government/publications/reported-norovirus-outbreaks-in-hospitals-suspected-and-laboratory-confirmed-outbreaks-in-2015/suspected-and-laboratory-confirmed-reported-norovirus-outbreaks-in-hospitals-outbreaks-occurring-in-weeks-49-to-53-2015. Accessed 22/09/2016.

99. Lopman BA, Reacher MH, Vipond IB, et al. Epidemiology and cost of nosocomial gastroenteritis, Avon, England, 2002-2003. Emerging infectious diseases **2004**; 10(10): 1827-34.

100. Sadique Z, Lopman B, Cooper BS, Edmunds WJ. Cost-effectiveness of Ward Closure to Control Outbreaks of Norovirus Infection in United Kingdom National Health Service Hospitals. The Journal of infectious diseases **2016**; 213 Suppl 1: S19-26.

101. Melhem NM. Norovirus vaccines: Correlates of protection, challenges and limitations. Human vaccines & immunotherapeutics **2016**; 12(7): 1653-69.

102. Atmar RL, Bernstein DI, Harro CD, et al. Norovirus vaccine against experimental human Norwalk Virus illness. The New England journal of medicine **2011**; 365(23): 2178-87.

103. Bernstein DI, Atmar RL, Lyon GM, et al. Norovirus vaccine against experimental human GII.4 virus illness: a challenge study in healthy adults. The Journal of infectious diseases **2015**; 211(6): 870-8.

104. Murray CJ, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet **2012**; 380(9859): 2197-223.

105. Lozano R, Naghavi M, Foreman K, et al. Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. Lancet **2012**; 380(9859): 2095-128.

106. Fischer Walker CL, Perin J, Aryee MJ, Boschi-Pinto C, Black RE. Diarrhea incidence in low- and middle-income countries in 1990 and 2010: a systematic review. BMC public health **2012**; 12: 220.

107. Ahmed SM, Hall AJ, Robinson AE, et al. Global prevalence of norovirus in cases of gastroenteritis: a systematic review and meta-analysis. The Lancet Infectious diseases **2014**; 14(8): 725-30.

108. de Wit MA, Koopmans MP, Kortbeek LM, et al. Sensor, a population-based cohort study on gastroenteritis in the Netherlands: incidence and etiology. American journal of epidemiology **2001**; 154(7): 666-74.

109. Scallan E, Hoekstra RM, Angulo FJ, et al. Foodborne illness acquired in the United States--major pathogens. Emerging infectious diseases **2011**; 17(1): 7-15.

110. Koo HL, Neill FH, Estes MK, et al. Noroviruses: The Most Common Pediatric Viral Enteric Pathogen at a Large University Hospital After Introduction of Rotavirus Vaccination. Journal of the Pediatric Infectious Diseases Society **2013**; 2(1): 57-60.

111. Payne DC, Vinje J, Szilagyi PG, et al. Norovirus and medically attended gastroenteritis in U.S. children. The New England journal of medicine **2013**; 368(12): 1121-30.

112. Nurminen K, Blazevic V, Huhti L, et al. Prevalence of norovirus GII-4 antibodies in Finnish children. Journal of medical virology **2011**; 83(3): 525-31.

113. Tam CC, Rodrigues LC, Viviani L, et al. Longitudinal study of infectious intestinal disease in the UK (IID2 study): incidence in the community and presenting to general practice. Gut **2012**; 61(1): 69-77.

114. Bartsch SM, Lopman BA, Ozawa S, Hall AJ, Lee BY. Global Economic Burden of Norovirus Gastroenteritis. PloS one **2016**; 11(4): e0151219.

115. van Kraaij MG, Dekker AW, Verdonck LF, et al. Infectious gastro-enteritis: an uncommon cause of diarrhoea in adult allogeneic and autologous stem cell transplant recipients. Bone marrow transplantation **2000**; 26(3): 299-303.

116. Schorn R, Hohne M, Meerbach A, et al. Chronic norovirus infection after kidney transplantation: molecular evidence for immune-driven viral evolution. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2010**; 51(3): 307-14.

117. Munir N, Liu P, Gastanaduy P, Montes J, Shane A, Moe C. Norovirus infection in immunocompromised children and children with hospital-acquired acute gastroenteritis. Journal of medical virology **2014**; 86(7): 1203-9.

118. Ye X, Van JN, Munoz FM, et al. Noroviruses as a Cause of Diarrhea in Immunocompromised Pediatric Hematopoietic Stem Cell and Solid Organ Transplant Recipients. Am J Transplant **2015**.

119. Lopman BA, Adak GK, Reacher MH, Brown DW. Two epidemiologic patterns of norovirus outbreaks: surveillance in England and wales, 1992-2000. Emerging infectious diseases **2003**; 9(1): 71-7.

120. Roddie C, Paul JP, Benjamin R, et al. Allogeneic hematopoietic stem cell transplantation and norovirus gastroenteritis: a previously unrecognized cause of morbidity. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2009**; 49(7): 1061-8.

121. Ahmed SM, Lopman BA, Levy K. A systematic review and meta-analysis of the global seasonality of norovirus. PloS one **2013**; 8(10): e75922.

122. Robilotti E, Deresinski S, Pinsky BA. Norovirus. Clinical microbiology reviews **2015**; 28(1): 134-64.

123. Debbink K, Donaldson EF, Lindesmith LC, Baric RS. Genetic mapping of a highly variable norovirus GII.4 blockade epitope: potential role in escape from human herd immunity. Journal of virology **2012**; 86(2): 1214-26.

124. Lindesmith LC, Beltramello M, Donaldson EF, et al. Immunogenetic mechanisms driving norovirus GII.4 antigenic variation. PLoS pathogens **2012**; 8(5): e1002705.

125. Allen DJ, Gray JJ, Gallimore CI, Xerry J, Iturriza-Gomara M. Analysis of amino acid variation in the P2 domain of the GII-4 norovirus VP1 protein reveals putative variant-specific epitopes. PloS one **2008**; 3(1): e1485.

126. Loveridge P, Cooper D, Elliot AJ, et al. Vomiting calls to NHS Direct provide an early warning of norovirus outbreaks in hospitals. The Journal of hospital infection **2010**; 74(4): 385-93.

127. Vinje J. Advances in Laboratory Methods for Detection and Typing of Norovirus. Journal of clinical microbiology **2014**.

128. Chen SY, Chiu CH. Worldwide molecular epidemiology of norovirus infection. Paediatrics and international child health **2012**; 32(3): 128-31.

129. Gallimore CI, Iturriza-Gomara M, Xerry J, Adigwe J, Gray JJ. Inter-seasonal diversity of norovirus genotypes: emergence and selection of virus variants. Archives of virology **2007**; 152(7): 1295-303.

130. Franck KT, Nielsen RT, Holzknecht BJ, Ersboll AK, Fischer TK, Bottiger B. Norovirus Genotypes in Hospital Settings: Differences Between Nosocomial and Community-Acquired Infections. The Journal of infectious diseases **2015**; 212(6): 881-8.

131. de Graaf M, van Beek J, Vennema H, et al. Emergence of a novel GII.17 norovirus - End of the GII.4 era? Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin **2015**; 20(26).

132. Chan MC, Lee N, Hung TN, et al. Rapid emergence and predominance of a broadly recognizing and fast-evolving norovirus GII.17 variant in late 2014. Nature communications **2015**; 6: 10061.

133. Verhoef L, Vennema H, van Pelt W, et al. Use of norovirus genotype profiles to differentiate origins of foodborne outbreaks. Emerging infectious diseases **2010**; 16(4): 617-24.

134. Verhoef L, Hewitt J, Barclay L, et al. Norovirus genotype profiles associated with foodborne transmission, 1999-2012. Emerging infectious diseases **2015**; 21(4): 592-9.

135. Vega E, Barclay L, Gregoricus N, Shirley SH, Lee D, Vinje J. Genotypic and epidemiologic trends of norovirus outbreaks in the United States, 2009 to 2013. Journal of clinical microbiology **2014**; 52(1): 147-55.

136. Gallimore CI, Pipkin C, Shrimpton H, et al. Detection of multiple enteric virus strains within a foodborne outbreak of gastroenteritis: an indication of the source of contamination. Epidemiology and infection **2005**; 133(1): 41-7.

137. Wang X, Yong W, Shi L, et al. An outbreak of multiple norovirus strains on a cruise ship in China, 2014. Journal of applied microbiology **2015**.

138. Hohne M, Niendorf S, Mas Marques A, Bock CT. Use of sequence analysis of the P2 domain for characterization of norovirus strains causing a large multistate outbreak of norovirus gastroenteritis in Germany 2012. International journal of medical microbiology : IJMM **2015**; 305(7): 612-8.

139. Muller L, Schultz AC, Fonager J, et al. Separate norovirus outbreaks linked to one source of imported frozen raspberries by molecular analysis, Denmark, 2010-2011. Epidemiology and infection **2015**; 143(11): 2299-307.

140. Sukhrie FH, Beersma MF, Wong A, et al. Using molecular epidemiology to trace transmission of nosocomial norovirus infection. Journal of clinical microbiology **2011**; 49(2): 602-6.

141. Kroneman A, Vennema H, Deforche K, et al. An automated genotyping tool for enteroviruses and noroviruses. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology **2011**; 51(2): 121-5.

142. Sukhrie FH, Teunis P, Vennema H, et al. P2 domain profiles and shedding dynamics in prospectively monitored norovirus outbreaks. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology **2013**; 56(4): 286-92.

143. Xerry J, Gallimore CI, Iturriza-Gomara M, Allen DJ, Gray JJ. Transmission events within outbreaks of gastroenteritis determined through analysis of nucleotide sequences of the P2 domain of genogroup II noroviruses. Journal of clinical microbiology **2008**; 46(3): 947-53.

144. Holzknecht BJ, Franck KT, Nielsen RT, Bottiger B, Fischer TK, Fonager J. Sequence analysis of the capsid gene during a genotype II.4 dominated norovirus season in one university hospital: identification of possible transmission routes. PloS one **2015**; 10(1): e0115331.

145. Verhoef L, Williams KP, Kroneman A, Sobral B, van Pelt W, Koopmans M. Selection of a phylogenetically informative region of the norovirus genome for outbreak linkage. Virus genes **2012**; 44(1): 8-18.

146. Eden JS, Tanaka MM, Boni MF, Rawlinson WD, White PA. Recombination within the pandemic norovirus GII.4 lineage. Journal of virology **2013**; 87(11): 6270-82.

147. Wong TH, Dearlove BL, Hedge J, et al. Whole genome sequencing and de novo assembly identifies Sydney-like variant noroviruses and recombinants during the winter 2012/2013 outbreak in England. Virology journal **2013**; 10: 335.

148. Kundu S, Lockwood J, Depledge DP, et al. Next-generation whole genome sequencing identifies the direction of norovirus transmission in linked patients. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2013**; 57(3): 407-14.

149. Cotten M, Petrova V, Phan MV, et al. Deep sequencing of norovirus genomes defines evolutionary patterns in an urban tropical setting. Journal of virology **2014**; 88(19): 11056-69.

150. Bull RA, Eden JS, Rawlinson WD, White PA. Rapid evolution of pandemic noroviruses of the GII.4 lineage. PLoS pathogens **2010**; 6(3): e1000831.

151. Kageyama T, Kojima S, Shinohara M, et al. Broadly reactive and highly sensitive assay for Norwalk-like viruses based on real-time quantitative reverse transcription-PCR. Journal of clinical microbiology **2003**; 41(4): 1548-57.

152. Stals A, Baert L, Botteldoorn N, et al. Multiplex real-time RT-PCR for simultaneous detection of GI/GII noroviruses and murine norovirus 1. Journal of virological methods **2009**; 161(2): 247-53.

153. Zeng SQ, Halkosalo A, Salminen M, Szakal ED, Puustinen L, Vesikari T. One-step quantitative RT-PCR for the detection of rotavirus in acute gastroenteritis. Journal of virological methods **2008**; 153(2): 238-40.

154. Oka T, Katayama K, Hansman GS, et al. Detection of human sapovirus by real-time reverse transcription-polymerase chain reaction. Journal of medical virology **2006**; 78(10): 1347-53.

155. Logan C, O'Leary JJ, O'Sullivan N. Real-time reverse transcription PCR detection of norovirus, sapovirus and astrovirus as causative agents of acute viral gastroenteritis. Journal of virological methods **2007**; 146(1-2): 36-44.

156. Bok K, Abente EJ, Realpe-Quintero M, et al. Evolutionary dynamics of GII.4 noroviruses over a 34-year period. Journal of virology **2009**; 83(22): 11890-901.

157. Kent WJ. BLAT--the BLAST-like alignment tool. Genome research **2002**; 12(4): 656-64.

158. Beaucourt S, Borderia AV, Coffey LL, et al. Isolation of fidelity variants of RNA viruses and characterization of virus mutation frequency. Journal of visualized experiments : JoVE **2011**; (52).

159. Tan M, Huang P, Meller J, Zhong W, Farkas T, Jiang X. Mutations within the P2 domain of norovirus capsid affect binding to human histo-blood group antigens: evidence for a binding pocket. Journal of virology **2003**; 77(23): 12562-71.

160. Lion T. Adenovirus infections in immunocompetent and immunocompromised patients. Clinical microbiology reviews **2014**; 27(3): 441-62.

161. Lion T, Kosulin K, Landlinger C, et al. Monitoring of adenovirus load in stool by real-time PCR permits early detection of impending invasive infection in patients after allogeneic stem cell transplantation. Leukemia **2010**; 24(4): 706-14.

162. Oka T, Wang Q, Katayama K, Saif LJ. Comprehensive review of human sapoviruses. Clinical microbiology reviews **2015**; 28(1): 32-53.

163. Dennehy PH. Rotavirus vaccines: an overview. Clinical microbiology reviews **2008**; 21(1): 198-208.

164. Atchison C, Collins S, Brown D, Ramsay ME, Ladhani S. Reduction in rotavirus disease due to the infant immunisation programme in England; evidence from national surveillance. The Journal of infection **2015**; 71(1): 128-31.

165. Markkula J, Hemming M, Vesikari T. Detection of vaccine-derived rotavirus strains in nonimmunocompromised children up to 3-6 months after RotaTeq(R) vaccination. The Pediatric infectious disease journal **2015**; 34(3): 296-8.

166. Bakare N, Menschik D, Tiernan R, Hua W, Martin D. Severe combined immunodeficiency (SCID) and rotavirus vaccination: reports to the Vaccine Adverse Events Reporting System (VAERS). Vaccine **2010**; 28(40): 6609-12.

167. Kaplon J, Cros G, Ambert-Balay K, et al. Rotavirus vaccine virus shedding, viremia and clearance in infants with severe combined immune deficiency. The Pediatric infectious disease journal **2015**; 34(3): 326-8.

168. Cunliffe NA, Booth JA, Elliot C, et al. Healthcare-associated viral gastroenteritis among children in a large pediatric hospital, United Kingdom. Emerging infectious diseases **2010**; 16(1): 55-62.

169. Lindell AT, Grillner L, Svensson L, Wirgart BZ. Molecular epidemiology of norovirus infections in Stockholm, Sweden, during the years 2000 to 2003: association of the GGIIb genetic cluster with infection in children. Journal of clinical microbiology **2005**; 43(3): 1086-92.

170. Yi J, Wahl K, Sederdahl BK, et al. Molecular epidemiology of norovirus in children and the elderly in Atlanta, Georgia, United States. Journal of medical virology **2015**.

171. Comans-Bitter WM, de Groot R, van den Beemd R, et al. Immunophenotyping of blood lymphocytes in childhood. Reference values for lymphocyte subpopulations. The Journal of pediatrics **1997**; 130(3): 388-93.

172. Gonzalez GG, Liprandi F, Ludert JE. Molecular epidemiology of enteric viruses in children with sporadic gastroenteritis in Valencia, Venezuela. Journal of medical virology **2011**; 83(11): 1972-82.

173. Franck KT, Fonager J, Ersboll AK, Bottiger B. Norovirus epidemiology in community and health care settings and association with patient age, Denmark. Emerging infectious diseases **2014**; 20(7): 1123-31.

174. Wheeler JG, Sethi D, Cowden JM, et al. Study of infectious intestinal disease in England: rates in the community, presenting to general practice, and reported to national surveillance. The Infectious Intestinal Disease Study Executive. BMJ (Clinical research ed) **1999**; 318(7190): 1046-50.

175. Rouhani S, Penataro Yori P, Paredes Olortegui M, et al. Norovirus Infection and Acquired Immunity in 8 Countries: Results From the MAL-ED Study. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2016**; 62(10): 1210-7.

176. Ye X, Van JN, Munoz FM, et al. Noroviruses as a Cause of Diarrhea in Immunocompromised Pediatric Hematopoietic Stem Cell and Solid Organ Transplant Recipients. Am J Transplant **2015**; 15(7): 1874-81.

177. Lee LY, Ison MG. Diarrhea caused by viruses in transplant recipients. Transplant infectious disease : an official journal of the Transplantation Society **2014**; 16(3): 347-58.

178. Batty EM, Wong TH, Trebes A, et al. A modified RNA-Seq approach for whole genome sequencing of RNA viruses from faecal and blood samples. PloS one **2013**; 8(6): e66129.

179. Nakamura S, Yang CS, Sakon N, et al. Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. PloS one **2009**; 4(1): e4219.

180. Depledge DP, Palser AL, Watson SJ, et al. Specific capture and whole-genome sequencing of viruses from clinical samples. PloS one **2011**; 6(11): e27805.

181. Liao GJ, Lun FM, Zheng YW, et al. Targeted massively parallel sequencing of maternal plasma DNA permits efficient and unbiased detection of fetal alleles. Clinical chemistry **2011**; 57(1): 92-101.

182. Thomson E, Ip CL, Badhan A, et al. Comparison of Next-Generation Sequencing Technologies for Comprehensive Assessment of Full-Length Hepatitis C Viral Genomes. **2016**; 54(10): 2470-84.

183. Bonsall D, Da Silva Filipe A, Hinds C, et al. Complete Genome Sequence of the Human Herpesvirus 6A Strain AJ from Africa Resembles Strain GS from North America. Journal of clinical microbiology **2015**; 3(1).

184. Donaldson CD, Clark DA, Kidd IM, Breuer J, Depledge DD. Genome Sequence of Human Herpesvirus 7 Strain UCL-1. Genome announcements **2013**; 1(5).

185. Brown AC, Bryant JM, Einer-Jensen K, et al. Rapid Whole-Genome Sequencing of Mycobacterium tuberculosis Isolates Directly from Clinical Samples. Journal of clinical microbiology **2015**; 53(7): 2230-7.

186. Christiansen MT, Brown AC, Kundu S, et al. Whole-genome enrichment and sequencing of Chlamydia trachomatis directly from clinical samples. BMC infectious diseases **2014**; 14: 591.

187. Dohm JC, Lottaz C, Borodina T, Himmelbauer H. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic acids research **2008**; 36(16): e105.

188. Brown JR, Gilmour K, Breuer J. Norovirus Infections Occur in B-Cell-Deficient Patients. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America **2016**; 62(9): 1136-8.

189. Bavelaar HH, Rahamat-Langendoen J, Niesters HG, Zoll J, Melchers WJ. Whole genome sequencing of fecal samples as a tool for the diagnosis and genetic characterization of norovirus. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology **2015**; 72: 122-5.

190. Chhabra P, Walimbe AM, Chitambar SD. Complete genome characterization of Genogroup II norovirus strains from India: Evidence of recombination in ORF2/3 overlap. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases **2010**; 10(7): 1101-9.

191. Won YJ, Park JW, Han SH, et al. Full-genomic analysis of a human norovirus recombinant GII.12/13 novel strain isolated from South Korea. PloS one **2013**; 8(12): e85063.

192. Botha JC, Taylor MB, Mans J. Comparative analysis of South African norovirus GII.4 strains identifies minor recombinant variants. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases **2016**.

193. Holmes EC, Ghedin E, Halpin RA, et al. Extensive geographical mixing of 2009 human H1N1 influenza A virus in a single university community. Journal of virology **2011**; 85(14): 6923-9.

194. Lopman BA, Gallimore C, Gray JJ, et al. Linking healthcare associated norovirus outbreaks: a molecular epidemiologic method for investigating transmission. BMC infectious diseases **2006**; 6: 108.

195. Xerry J, Gallimore CI, Cubitt D, Gray JJ. Tracking environmental norovirus contamination in a pediatric primary immunodeficiency unit. Journal of clinical microbiology **2010**; 48(7): 2552-6.

196. Martella V, Medici MC, De Grazia S, et al. Evidence for recombination between pandemic GII.4 norovirus strains New Orleans 2009 and Sydney 2012. Journal of clinical microbiology **2013**; 51(11): 3855-7.

197. Fonager J, Barzinci S, Fischer TK. Emergence of a new recombinant Sydney 2012 norovirus variant in Denmark, 26 December 2012 to 22 March 2013. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin **2013**; 18(25).

198. Bruggink L, Catton M, Marshall J. A norovirus intervariant GII.4 recombinant in Victoria, Australia, June 2016: the next epidemic variant? Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin **2016**; 21(39).

199. Fonager J, Rasmussen LD, Fischer TK. Letter to the editor: A norovirus intervariant GII.4 recombinant in Victoria, Australia, June 2016: the next epidemic variant? Reflections and a note of caution. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin **2016**; 21(41).

200. Matsushima Y, Shimizu T, Ishikawa M, et al. Complete Genome Sequence of a Recombinant GII.P16-GII.4 Norovirus Detected in Kawasaki City, Japan, in 2016. Genome announcements **2016**; 4(5).

201. Mitchell C, Meredith P, Richardson M, Greengross P, Smith GB. Reducing the number and impact of outbreaks of nosocomial viral gastroenteritis: time-series analysis of a

multidimensional quality improvement initiative. BMJ quality & safety **2016**; 25(6): 466-74.

202. Evans MR, Meldrum R, Lane W, et al. An outbreak of viral gastroenteritis following environmental contamination at a concert hall. Epidemiology and infection **2002**; 129(2): 355-60.

203. Wu HM, Fornek M, Schwab KJ, et al. A norovirus outbreak at a long-term-care facility: the role of environmental surface contamination. Infection control and hospital epidemiology **2005**; 26(10): 802-10.

204. Tuladhar E, Hazeleger WC, Koopmans M, Zwietering MH, Duizer E, Beumer RR. Transfer of noroviruses between fingers and fomites and food products. International journal of food microbiology **2013**; 167(3): 346-52.

205. Kaufman SS, Chatterjee NK, Fuschino ME, et al. Characteristics of human calicivirus enteritis in intestinal transplant recipients. Journal of pediatric gastroenterology and nutrition **2005**; 40(3): 328-33.

206. Ghosh N, Malik FA, Daver RG, Vanichanan J, Okhuysen PC. Viral associated diarrhea in immunocompromised and cancer patients at a large comprehensive cancer center: a 10-year retrospective study. Infectious diseases (London, England) **2016**: 1-7.

207. Westhoff TH, Vergoulidou M, Loddenkemper C, et al. Chronic norovirus infection in renal transplant recipients. Nephrology, dialysis, transplantation : official publication of the European Dialysis and Transplant Association - European Renal Association **2009**; 24(3): 1051-3.

208. Simon A, Schildgen O, Maria Eis-Hubinger A, et al. Norovirus outbreak in a pediatric oncology unit. Scandinavian journal of gastroenterology **2006**; 41(6): 693-9.

209. Lee BE, Pang XL, Robinson JL, Bigam D, Monroe SS, Preiksaitis JK. Chronic norovirus and adenovirus infection in a solid organ transplant recipient. The Pediatric infectious disease journal **2008**; 27(4): 360-2.

210. Coste JF, Vuiblet V, Moustapha B, et al. Microbiological diagnosis of severe diarrhea in kidney transplant recipients by use of multiplex PCR assays. Journal of clinical microbiology **2013**; 51(6): 1841-9.

211. Ebdrup L, Bottiger B, Molgaard H, Laursen AL. Devastating diarrhoea in a heart-transplanted patient. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology **2011**; 50(4): 263-5.

212. Nilsson M, Hedlund KO, Thorhagen M, et al. Evolution of human calicivirus RNA in vivo: accumulation of mutations in the protruding P2 domain of the capsid leads to structural changes and possibly a new phenotype. Journal of virology **2003**; 77(24): 13117-24.

213. Engelen MA, Gunia S, Stypmann J. Elimination of norovirus in a chronic carrier under immunosuppression after heart transplantation--effect of everolimus. Transplant international : official journal of the European Society for Organ Transplantation **2011**; 24(11): e102-3.

214. Wingfield T, Gallimore CI, Xerry J, et al. Chronic norovirus infection in an HIV-positive patient with persistent diarrhoea: a novel cause. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology **2010**; 49(3): 219-22.

215. Woodward JM, Gkrania-Klotsas E, Cordero-Ng AY, et al. The role of chronic norovirus infection in the enteropathy associated with common variable immunodeficiency. The American journal of gastroenterology **2015**; 110(2): 320-7.

216. Duraisingham SS, Manson A, Grigoriadou S, Buckland M, Tong CY, Longhurst HJ. Immune deficiency: changing spectrum of pathogens. Clinical and experimental immunology **2015**; 181(2): 267-74.

217. van de Ven AA, Douma JW, Rademaker C, et al. Pleconaril-resistant chronic parechovirus-associated enteropathy in agammaglobulinaemia. Antiviral therapy **2011**; 16(4): 611-4.

218. Lee LY, Ladner DP, Ison MG. Norovirus infection in solid organ transplant recipients: a single-center retrospective study. Transplant infectious disease : an official journal of the Transplantation Society **2016**.

219. Trivedi TK, Desai R, Hall AJ, Patel M, Parashar UD, Lopman BA. Clinical characteristics of norovirus-associated deaths: a systematic literature review. American journal of infection control **2013**; 41(7): 654-7.

220. Kaufman SS, Chatterjee NK, Fuschino ME, et al. Calicivirus enteritis in an intestinal transplant recipient. Am J Transplant **2003**; 3(6): 764-8.

221. Echenique IA, Stosor V, Gallon L, Kaufman D, Qi C, Zembower TR. Prolonged norovirus infection after pancreas transplantation: a case report and review of chronic norovirus. Transplant infectious disease : an official journal of the Transplantation Society **2016**; 18(1): 98-104.

222. Florescu DF, Hill LA, McCartan MA, Grant W. Two cases of Norwalk virus enteritis following small bowel transplantation treated with oral human serum immunoglobulin. Pediatric transplantation **2008**; 12(3): 372-5.

223. Chagla Z, Quirt J, Woodward K, Neary J, Rutherford C. Chronic norovirus infection in a transplant patient successfully treated with enterally administered immune globulin. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology **2013**; 58(1): 306-8.

224. Gairard-Dory AC, Degot T, Hirschi S, et al. Clinical usefulness of oral immunoglobulins in lung transplant recipients with norovirus gastroenteritis: a case series. Transplantation proceedings **2014**; 46(10): 3603-5.

225. Boillat Blanco N, Kuonen R, Bellini C, et al. Chronic norovirus gastroenteritis in a double hematopoietic stem cell and lung transplant recipient. Transplant infectious disease : an official journal of the Transplantation Society **2011**; 13(2): 213-5.

226. Bruggink LD, Hayes L, Marshall JA. Chronic norovirus infection in a patient with a past history of Burkitt lymphoma. Virusdisease **2015**; 26(3): 207-10.

227. Ronchetti AM, Henry B, Ambert-Balay K, et al. Norovirus-related chronic diarrhea in a patient treated with alemtuzumab for chronic lymphocytic leukemia. BMC infectious diseases **2014**; 14: 239.

228. Knoll BM, Lindesmith LC, Yount BL, Baric RS, Marty FM. Resolution of diarrhea in an immunocompromised patient with chronic norovirus gastroenteritis correlates with constitution of specific antibody blockade titer. Infection **2016**; 44(4): 551-4.

229. Gallimore CI, Lewis D, Taylor C, Cant A, Gennery A, Gray JJ. Chronic excretion of a norovirus in a child with cartilage hair hypoplasia (CHH). Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology **2004**; 30(2): 196-204.

230. Chehade H, Girardin E, Delich V, Pascual MA, Venetz JP, Cachat F. Acute norovirus-induced agranulocytosis in a pediatric kidney transplant recipient. Transplant infectious disease : an official journal of the Transplantation Society **2012**; 14(4): E27-9.

231. Avery RK, Lonze BE, Kraus ES, Marr KA, Montgomery RA. Severe chronic norovirus diarrheal disease in transplant recipients: Clinical features of an under-recognized syndrome. Transplant infectious disease : an official journal of the Transplantation Society **2017**; 19(2).

232. Kaufman SS, Green KY, Korba BE. Treatment of norovirus infections: moving antivirals from the bench to the bedside. Antiviral research **2014**; 105: 80-91.

233. Drysdale SB, Green CA, Sande CJ. Best practice in the prevention and management of paediatric respiratory syncytial virus infection. Therapeutic advances in infectious disease **2016**; 3(2): 63-71.

234. Prusiner P, Sundaralingam M. A new class of synthetic nucleoside analogues with broad-spectrum antiviral properties. Nature: New biology **1973**; 244(134): 116-8.

235. Sidwell RW, Huffman JH, Khare GP, Allen LB, Witkowski JT, Robins RK. Broad-spectrum antiviral activity of Virazole: 1-beta-D-ribofuranosyl-1,2,4-triazole-3-carboxamide. Science (New York, NY) **1972**; 177(4050): 705-6.

236. Witkowski JT, Robins RK, Sidwell RW, Simon LN. Design, synthesis, and broad spectrum antiviral activity of 1- -D-ribofuranosyl-1,2,4-triazole-3-carboxamide and related nucleosides. Journal of medicinal chemistry **1972**; 15(11): 1150-4.

237. Vignuzzi M, Stone JK, Andino R. Ribavirin and lethal mutagenesis of poliovirus: molecular mechanisms, resistance and biological implications. Virus research **2005**; 107(2): 173-81.

238. Chang KO, George DW. Interferons and ribavirin effectively inhibit Norwalk virus replication in replicon-bearing cells. Journal of virology **2007**; 81(22): 12111-8.

239. Graci JD, Cameron CE. Mechanisms of action of ribavirin against distinct viruses. Reviews in medical virology **2006**; 16(1): 37-48.

240. Swetina J, Schuster P. Self-replication with errors. A model for polynucleotide replication. Biophysical chemistry **1982**; 16(4): 329-45.

241. Domingo E. Viruses at the edge of adaptation. Virology **2000**; 270(2): 251-3.

242. Eigen M. Error catastrophe and antiviral strategy. Proceedings of the National Academy of Sciences of the United States of America **2002**; 99(21): 13374-6.

243. Arias A, Thorne L, Ghurburrun E, Bailey D. Norovirus Polymerase Fidelity Contributes to Viral Transmission In Vivo. **2016**; 1(5).

244. Arias A, Thorne L, Goodfellow I. Favipiravir elicits antiviral mutagenesis during virus replication in vivo. eLife **2014**; 3: e03679.

245. Florescu DF, Hermsen ED, Kwon JY, et al. Is there a role for oral human immunoglobulin in the treatment for norovirus enteritis in immunocompromised patients? Pediatric transplantation **2011**; 15(7): 718-21.

246. Siddiq DM, Koo HL, Adachi JA, Viola GM. Norovirus gastroenteritis successfully treated with nitazoxanide. The Journal of infection **2011**; 63(5): 394-7.

247. Rossignol JF, El-Gohary YM. Nitazoxanide in the treatment of viral gastroenteritis: a randomized double-blind placebo-controlled clinical trial. Alimentary pharmacology & therapeutics **2006**; 24(10): 1423-30.

248. Victoria M, Tort LF, Lizasoain A, et al. Norovirus molecular detection in Uruguayan sewage samples reveals a high genetic diversity and GII.4 variant replacement along time. Journal of applied microbiology **2016**; 120(5): 1427-35.

249. Mans J, Netshikweta R, Magwalivha M, Van Zyl WB, Taylor MB. Diverse norovirus genotypes identified in sewage-polluted river water in South Africa. Epidemiology and infection **2013**; 141(2): 303-13.

250. van den Berg H, Lodder W, van der Poel W, Vennema H, de Roda Husman AM. Genetic diversity of noroviruses in raw and treated sewage water. Research in microbiology **2005**; 156(4): 532-40.

251. Karst SM, Baric RS. What is the reservoir of emergent human norovirus strains? Journal of virology **2015**; 89(11): 5756-9.

252. Grenfell BT, Pybus OG, Gog JR, et al. Unifying the epidemiological and evolutionary dynamics of pathogens. Science (New York, NY) **2004**; 303(5656): 327-32.

253. Simon AK, Hollander GA, McMichael A. Evolution of the immune system in humans from infancy to old age. Proceedings Biological sciences / The Royal Society **2015**; 282(1821): 20143085.

254. Hobbs MA, Robinson S, Neyens DM, Steed C. Visitor characteristics and alcohol-based hand sanitizer dispenser locations at the hospital entrance: Effect on visitor use rates. American journal of infection control **2016**; 44(3): 258-62.

255. Birnbach DJ, Nevo I, Barnes S, et al. Do hospital visitors wash their hands? Assessing the use of alcohol-based hand sanitizer in a hospital lobby. American journal of infection control **2012**; 40(4): 340-3.

256. Vaidotas M, Yokota PK, Marra AR, et al. Measuring hand hygiene compliance rates at hospital entrances. American journal of infection control **2015**; 43(7): 694-6.

257. Randle J, Firth J, Vaughan N. An observational study of hand hygiene compliance in paediatric wards. Journal of clinical nursing **2013**; 22(17-18): 2586-92.

258. Snow M, White GL, Jr., Kim HS. Inexpensive and time-efficient hand hygiene interventions increase elementary school children's hand hygiene rates. The Journal of school health **2008**; 78(4): 230-3.

259. Wang CM, Chen SC, Chen KT. Current status of rotavirus vaccines. World journal of pediatrics : WJP **2015**; 11(4): 300-8.